Dynamics of co-existing Escherichia colilineages in situ of the infant gut and multiplex phenotypic targeted recovery of previously uncultivated bacteria from the human gut

Gumpert, Heidi

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Ph.D. Thesis
Doctor of Philosophy

DTU Systems Biology
Department of Systems Biology

Dynamics of co-existing *Escherichia coli* lineages *in situ* of the infant gut

and multiplex phenotypic targeted recovery of previously uncultivated bacteria from the human gut

Heidi Kira Gumpert

Kongens Lyngby 2014

DTU Systems Microbiology
Centre for Systems Microbiology
Technical University of Denmark

Matematiktorvet
Building 301
2800 Kongens Lyngby, Denmark
www.bio.dtu.dk

Abstract

The work in this thesis explores the dynamic nature of *Escherichia coli* lineages co-existing in the human intestinal tract. The work is supported via full genome sequencing of co-existing *E. coli* strains isolated from infants enrolled in the ALLERGYFLORA study. Both sets of isolates examined here were selected due to an observed change in their antibiotic susceptibility profile. Via full genome sequencing, we identified that in both cases a conjugative plasmid harboring antibiotic resistance genes was transferred between co-existing *E. coli* lineages and is responsible for the change in antibiotic susceptibility. In one case, the transfer occurred in the absence of antibiotic treatment and the transconjugant remained amongst the gut microbiota for months, providing evidence to the hypothesis that resistance genes are stably maintained once acquired. To our knowledge, this is the first documented case where transfer of resistance genes was observed *in situ* of a non-perturbed gut in the absence of antibiotic treatment. Additionally, cases of genomic deletions, phage infections and curing, and plasmid loss were observed, highlighting that strains in the gut are highly dynamic. The detection of these micro-evolutionary events was facilitated through the isolation of strains, yielding a high-resolution view into the strain heterogeneity of the human gut.

Also included in this thesis is work highlighting that a large proportion of the gut microbiota can indeed be cultivated using carefully designed conditions. Antibiotic tolerance phenotypes were determined, and this mapping was used to carefully tailor antibiotic combinations to specifically select for previously uncultivated bacteria. Using this method, four previously uncultivated species were successfully cultured and genome sequenced, and two of which had 16S rRNA identities of less than 95% to previously cultured bacteria. We assessed the genomic coverage and abundance of these sequenced isolates in the gut using publicly available metagenomes.

Dansk Resume

Formålet med denne afhandling er at afdække de dynamiske egenskaber mellem flere stammer af *Escherichia coli* ved deres sameksistens i den menneskelige tarmkanal. Arbejdet er baseret på genomisk sekventering af sameksisterende *E. coli* stammer isoleret fra nyfødte, der undergik "ALLERGYFLORA" undersøgelsen. De to set prøver blev udvalgt på baggrund af en observeret forandring i deres modtagelighed over for antibiotika. Via genomisk sekventering identificerede vi i begge tilfælde et plasmid med antibiotisk resistente gener, var blevet overført mellem forskellige stammer af *E. coli*, og derved medføre en ændring i stammernes modtagelighed for antibiotika. I et tilfælde foregik overførelsen af gener i fravær af antibiotika, hvorefter modtageren af generne forblev i fordøjelses systemet i månedsvis. Derved underbygger det tesen om, at gener kan regnes for stabile, når de først er tilegnede. Det er til vores viden første gang overførslen af resistente gener er observerede naturligt i tarmsystemet, i fravær af antibiotika. Ydermere har vi også observeret, genomiske deletioner, virus infektioner og tabet af plasmider. Dette underbygger vores tese om, at tarmsystemet er et meget dynamisk sted, på bakteriernes genomiske stamme niveau.

Preface

This PhD thesis was prepared at the Centre for Systems Microbiology, in the Department of Systems Biology at the Technical University of Denmark in fulfillment of the requirements for the PhD degree.

Kongens Lyngby, October 15, 2014

Heidi Kira Gumpert

Acknowledgements

I would first like to thank everyone who has been a part of the Sommer lab during my time there. It has been both a pleasure to work with you, and to get to know you all. Thanks for all the discussions, and thanks for all the coffees and cake.

My profound gratitude to everyone who has provided me with guidance in the lab and other help along the way. Thank you first to Morten, for your supervision, guidance and motivation throughout my time in your group. From Mari helping me out at the very beginning (and throughout), and for keeping the lab in good working order. To Christian, my main partner in crime, I thank you for all the work and discussion of the projects we've done together. And thanks to my other fellow PhDs, Dionisio and Rachel, as well. Thank you to my favourite phage girl, Lejla, for both your discussions and advice, and your keen wit. Elizabeth, I have really appreciated sharing my fondness over gut microbes with you, and you have been perhaps the only one whom I could discuss working with our type of sample matter with. Thank you to Andreas for being a super helpful and knowledgeable lab bench neighbour - best of luck to you now starting your PhD.

A special thanks to all my office mates over the years, from my most recent 'homeboys' Marius and Mostafa, and to Thomas, and the full house of Marie, Kira, Hans and Mads. You guys had the honour of being closest to my particular brand of wonderment - thanks for that. And thanks for adding to a healthy office milieu.

I would also like to thank Nahid Karami, Agnes Wold and Ingegerd Alderberth from the Sahlgrenska Academy, our collaborators on the *E. coli* colonization in Swedish infants project. Thanks for working with Morten and I, and thank you for the discussions we've had and your input on the project.

And of course warmest thanks to all the dear friends and family in my life.

Abbreviations

ABU	asymptomatic bacteriuria
AIEC	adhesive invasive *E. coli*
Amp	ampicillin
bp	base pairs
CF	cystic fibrosis
CFU	colony forming unit
DGGE	denaturing gradient gel electrophoresis
DM25	Davis medium with 25mg/mL glucose
EIEC	enteroinvasive *E. coli*
ESBL	extended spectrum beta-lactamase
FISH	fluorescent *in situ* hybridization
HGT	horizontal gene transfer
HMP	Human microbiome project
HUS	hemolytic-uremic syndrome
INDEL	insertion and/or deletion
IS	insertion sequence
kb	kilobases
MDR	multi-drug resistant
MHBII	Mueller Hinton broth II, cation-adjusted
MIC	minimum inhibitory concentration
MLST	multi locus sequence typing
MRSA	methicillin resistant *Staphylococcus aureus*
NMEC	neotal meningitis *E. coli*
OD	optical density
ORF	open reading frame
PCR	polymerase chain reaction
PE	paired-end (reads)
PFGE	pulsed-field gel electrophoresis
Pip	piperacillin
RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism

Abbreviations, continued.

SD	standard deviation
SE	single-end (reads)
Sfx	sulfamethoxazole
SNP	single nucleotide polymorphism
Str	streptomycin
TB	tuberculosis
TGGE	temperature gel gradient electrophoresis
T-RFLP	terminal restriction fragment length polymorphism
UPEC	uropathogenic *E. coli*
UTI	urinary tract infection
VRE	vancomycin resistant *Enterococcus*

Contents

CHAPTER 1

Outline of Thesis

This thesis has been organised into a total of six chapters. This first chapter outlines the rest of the subsequent chapters. Chapter two includes a list of publications that the author contributed to, highlighting which publications are included in this thesis.

Chapter three provides an introduction to the subject areas of antibiotic resistance, horizontal gene transfer and reservoirs of antibiotic resistance genes, and the human gut microbiota. Relevant work is also reviewed here, including how the human gut microbiota is studied, dynamics of the lineages of the human gut microbiota and horizontal gene exchange within the human gut.

Chapters four and five of this thesis discuss two cases of horizontal transfer of antibiotic resistance genes between co-existing *E. coli* lineages in the gut. The first case occurs in the absence of antibiotic treatment, and the transconjugant is even maintained in the gut microbiota for months after the transconjugant was detected. The second case occurs during antibiotic treatment, and the transconjugant is not detected at a later time point. The analysis of these sets of isolates showed that these strains existing in the gut are highly dynamic, and multiple cases of phage infections and clearance, plasmid acquisition and loss, and large genomic deletions are reported.

The final chapter of this thesis discusses how selecting carefully designed culture conditions a large proportion of the gut microbiota can indeed be cultivated as opposed to current belief. Specifically, antibiotic tolerance phenotypes were determined, and this mapping was used to carefully tailor antibiotic combinations to specifically select for previously uncultivated bacteria. Using this method, four previously uncultivated species were successfully cultured and genome sequenced.

CHAPTER 2

List of Publications

Included in this thesis

H.K. Gumpert*, N. Karami*, C. Munck, A.E. Wold, I. Adlerberth, M.O.A. Sommer. Transfer of multiple antibiotic resistance genes in situ of the infant gut microbiota. *Manuscript not submitted.*

H.K. Gumpert, N. Karami, A.E. Wold, I. Adlerberth, M.O.A. Sommer. Dynamics of pathogenic and commensal Escherichia coli lineages in situ of the gut during antibiotic treatment. *Manuscript not submitted.*

E A. Rettedal, **H.K. Gumpert** and M.O.A. Sommer. Cultivation-based multiplex phenotyping of human gut microbiota allows targeted recovery of previously uncultured bacteria. *Nature Communications*, 5:4714, August 2014.

Other publications

C. Munck, **H.K. Gumpert**, A.I. Nilsson Wallin, H.Wang and M.O.A. Sommer. Resistance development against drug combinations is predicted by the collateral responses to the component drugs. *Undergoing final corrections for Science Translational Medicine.*

M. Rodriguez de Evgrafova, **H.K. Gumpert**, C. Munck, T.T. Thomsen, and M.O.A. Sommer. Collateral resistance and sensitivity modulate evolution of high-level resistance to drug combination treatment in *Staphylococcus aureus*. *Undergoing final corrections for Molecular Biology and Evolution.*

CHAPTER 3

Introduction

Antibiotic resistance is an ever-growing public health concern. Antibiotic resistance, where bacteria are able to resist the effects of an antibiotic they were once susceptible to, results in reduced efficacy of the antibiotic and may cause treatment failure, leading to prolonged illness and increased mortality [1]. The search for new antibiotics to treat infections that are no longer susceptible to previously used antibiotics is ongoing. However, with a long innovation gap and other difficulties, there have been very few new antibiotic drugs brought to market [2]. The World Health Organization has recently reported that the threat of the post antibiotic era, where the current antibiotics are no longer able to treat common infections, may indeed be a real possibility for the 21st century [1]. While one part of the solution to the current antibiotic resistance problem includes the discovery of new antibiotics, much focus is directed on solutions to curb the increase of antibiotic resistant bacteria.

3.1 Antibiotic Resistance Mechanisms

Antibiotic resistance is achieved by a variety of tactics (Fig. 3.1). Resistance may be achieved through reduced uptake or through active elimination of the drug, both resulting in a reduced and insufficient drug concentration inside the bacterium to be effective. Reduced drug uptake via low permeability can be due to intrinsic drug resistance, where the bacterium is not susceptible to the drug due to its physiology, but it can also be due to a change or loss of uptake proteins such as porins [3]. Efflux pumps, which actively export compounds, are widespread amongst bacteria and have been found to export a broad range of antibiotics and other biocides [4]. Resistance can also be achieved via an alteration of the drug target no longer allowing the drug to interact with the target in an effective way. For example, resistance to the quinolone drug, nalidixic acid, can be achieved in *Escherichia coli* via mutation altering the interaction between DNA gyrase and nalidixic acid, such that it can no longer bind in an efficient manner [5, 6]. Alternatively the drug target can also be modified post-

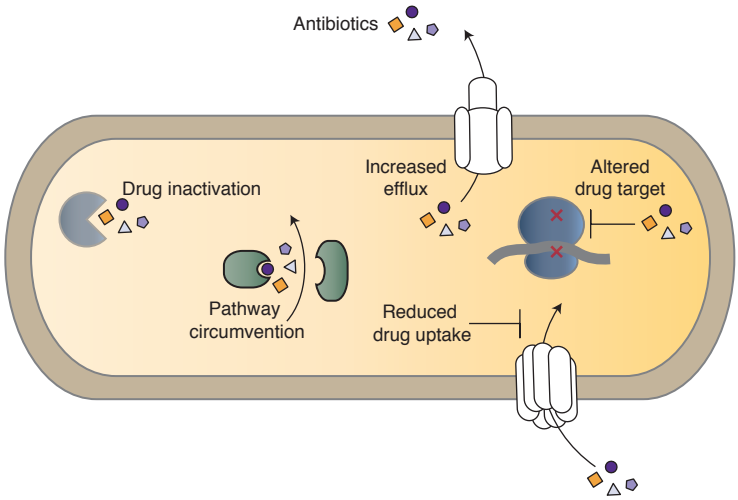


Figure 3.1 | Mechanisms of antibiotic resistance. Figure courtesy of Christian Munck.

transcription, as in the case of macrolide resistance involving post-transcriptional methylation of the 23S ribosome [7]. Resistance can also arise via the ability to break down the drug, rendering it inactive. Beta-lactamase enzymes illustrate a classic example of resistance achieved through degradation of the drug. These enzymes are able to cleave the beta-lactam ring of the drug, resulting in inactivation [8]. Other mechanisms also exist such as bypassing the antibiotic inhibited reaction, sequestering the antibiotic via protein binding, or titration of the antibiotic by increasing the number of drug target [8]. Additionally, it is not uncommon that more than one of these mechanisms can confer resistance to a single antibiotic, and for one mechanism to confer resistance to multiple antibiotics.

3.2 Routes to Antibiotic Resistance

A bacterium can become resistant to an antibiotic via two routes. It can be achieved vertically through inheritance or horizontally via acquisition of foreign genetic material that confers resistance [9]. Vertical transfer of antibiotic resistance involves

propagation of a particular resistant strain [10]. Vertical inheritance of antibiotic resistance determinants includes both resistances that have come about through mutation as well as through the acquisition of foreign genetic material. Horizontal gene transfer (HGT) of antibiotic resistance genes can occur through a variety of mechanisms, including phage transduction, conjugation or uptake of foreign DNA by naturally competent bacteria [9]. The origin of the genes conferring resistance need not be from a related species. Many plasmids have been found to have a wide host range, and even conjugative transfer between gram-positive and gram-negative bacteria have been reported [11].

3.3 Horizontal Gene Transfer Mechanisms

There are a variety of mechanisms for horizontal gene transfer. As previously mentioned, these mechanisms include phage transduction, conjugation or uptake of foreign DNA.

Phage transduction is the process by which a bacterial virus, a bacteriophage, includes a portion of its infected host's DNA into the virion and transfers that genetic information to another bacterium [12]. Once a phage has successfully entered the host bacterium, it will either undergo multiplication of the phage and subsequent lysis or it will incorporate its genetic material into the host's chromosome [13]. These options are called the lytic and lysogenic pathways, respectively. A phage entering the lysogenic state results in the host bacterium gaining additional genetic information. The phage may contain additional genes, such as antibiotic resistance genes, in addition to the phage-specific genes. A lysogenic phage may enter the lytic state, thereby being released to infect other bacteria. During this process, the phage may excise itself from the host's genome restoring the region to its uninfected state, or it may be excised with genomic material from the host. Phage transduction can either be specialized, where only DNA adjacent to the insertion site is accidentally packaged into the phage, or it can be generalized, where any DNA can be packaged into the phage [12].

Another method of horizontal gene transfer is via conjugation. In conjugation, bacteria are temporarily united in order to transfer genetic information. DNA is believed to be transferred via the hollow conjugation pilus, yet there is some debate whether the pili may play a role in the stabilization of the cells and that DNA may be transferred through a pore [14]. However, conjugation has been directly observed

between single *E. coli* cells and at considerable distances [15], and DNA has been detected within the pilus [16]. Single-stranded DNA is transferred to the receiver by separating the two DNA strands before it is passed through the conjugation pilus [17]. Conjugation is conservative for plasmids, as this splitting process leaves both the donor and a receiver with a template of the plasmid.

Conjugative systems exist both on plasmids and in conjugative transposons to facilitate mobilization of their genetic information. Generally, the vector encoding the conjugation machinery is transferred. However, other genetic elements that do not encode their own conjugation machinery can be transferred instead. For example, mobilizable plasmids carry only the genetic information for the processing of conjugative DNA, and rely on another element in the cell to provide the system to form a mating channel [18].

Lastly, bacteria that are naturally competent may take up naked DNA. This process is called transformation. Many naturally competent bacteria become competent in response to an environmental cue, such as cell density, nutrient access or limitation [19]. In contrast to the other two mechanisms of horizontal gene transfer, this mechanism is only dependent on the receiver taking up the external DNA and not on mobilization elements [20]. However, natural transformation still relies on exposure to extracellular DNA, which is continually released into the environment from deteriorated cells and viral particles, and from actively excreting cells [19]. Interestingly, many genera of bacteria have been found to actively excrete DNA [19]. The DNA that is transformed is typically short in length, approximately one to three genes, and is therefore less likely to be subjected to restriction enzyme digestion in the host [21].

Importantly, in all of these methods of horizontal gene transfer, either autonomous replication or integration into the host chromosome is necessary for continued propagation of the genetic information. Plasmids normally ensure their replication and correct segregation into daughter cells during cell division. If the plasmid relies on proteins to initiate replication, the replication proteins must be compatible with the host bacterium in order for successful replication of the plasmid [19]. Additionally, other plasmids using the same replication and partitioning systems are unable to co-exist in the same host [21]. For this reason, plasmids are often classified into 'incompatibility' groups based on possible co-existence. Furthermore, autonomous plasmids must ensure segregational stability, which is often achieved through a relatively high copy number of the plasmid and relying on segregation by chance, or through low copy number and active partitioning mechanisms [19]. Similar to transposons, phages, and

exogenous DNA, parts of plasmids may also be integrated into the chromosome [22]. Integration has different host demands. Integration can occur through homologous recombination, therefore being dependent on homology to the host [21]. Integration may occur through insertion sequence (IS) elements, requiring their presence on the chromosome, but independent of sequence homology [21].

3.4 Mutational vs. Horizontally Acquired Antibiotic Resistance

While the exact clinical contributions of either mutational- or horizontally-acquired antibiotic resistances are unclear, examples highlighting the significance of both exist [23]. Antibiotic resistance that arises through mutation is of significant clinical importance during long-term treatment regimes [24, 25]. One of the first infections to be treated via antibiotic chemotherapy was tuberculosis (TB) caused by *Mycobacterium tuberculosis*. Initially, streptomycin monotherapy treatment took a total of 24 months. This long exposure to antibiotics during treatment was met with emerging drug resistances. The role mutational resistances played were recognized, and drug-combination therapy was introduced. Although treatment time was reduced to 6 months, failures in adherence of treatment plans had unfortunately resulted in the development and global spread of multi-drug resistance *M. tuberculosis* [26, 27]. The significance of antibiotic resistance via mutation in TB is also due to the compartmental nature of the infection [24, 9]. This compartmentalization leaves *M. tuberculosis* little opportunity to interact with other strains and potentially acquiring resistance genes horizontally [24].

Pseudomonas aeruginosa is an opportunistic pathogen, and is of particular concern for cystic fibrosis (CF) patients. It causes a chronic airway infection in most CF patients, and ultimately most patients succumb to respiratory failure caused in conjunction with chronic bacterial infection and accompanying airway failure [28]. Patients often undergo long-term antibiotic treatment in an attempt to reduce the severity and improve long-term prognosis of the patients [29], but this long-term treatment is associated with the development of resistance [9]. Analysis of clinical *P. aeruginosa* isolates collected over a 35-year time period from CF patients have revealed that antibiotic resistance via the acquisition of genetic material does not play a role in resistance of *P. aeruginosa* for CF patients, and that it is rather acheived via mutation [30, 25, 31]. Specific examples of resistances found in *P. aeruginosa* iso-

lates from the lungs of CF patients include *gyrA* mutations conferring fluoroquinolone resistance [25], and de-repressed beta-lactamase production conferring beta-lactam resistance [32]. Therefore, for some bacterial infections, resistance via mutation plays a significant role due to the nature of the infection or the treatment regime.

For many human pathogens, horizontal gene transfer is significant in the development of multidrug resistant human pathogens [33].

The increase in multi-drug-resistant (MDR) *Acinetobacter baumannii* isolates from clinics around the world is a worrisome development, especially as some strains have been reported to be resistant to all known antibiotics [34, 35]. *Acinetobacter* spp. are among a unique class of gram-negative bacteria that are "naturally competent", and some strains have been shown to readily undergo recombination [34]. A genome sequencing study of an *A. baumannii* isolate identified that 17% of all open reading frames were contained in as many as 28 alien genomic islands, 8 of which contained antibiotic resistance genes, indicating that the genome has acquired a large amount of foreign DNA [36]. Genome analysis of a MDR *A. baumannii* epidemic strain revealed an 86kb resistance island with 45 identified resistance genes [35]. Furthermore, with 82 out of the 88 open reading frames predicted to have originated from other gram-negative organisms, such as *Pseudomonas* sp., *Salmonella* sp., and *E. coli* [35]. While resistance genes on conjugative plasmids have been the determinants of resistance for some *Acinetobacter* spp., it is currently transposons that have been found to play a significant role in the dissemination of antibiotic resistance determinants, particularly in the MDR phenotype [34].

In gram-negative *Enterobacteriaceae*, conjugative plasmids play a key role in the evolution and dissemination of beta-lactamase resistance genes in both nosocomial and community acquired infections [9]. A particularly striking example of this is the dissemination of the CTX-M genes, which are now globally dispersed and are the dominant type of extended spectrum beta-lactamases (ESBL) in Europe [33].

Amongst gram-positive bacteria, it is the commensal bacterium *Staphylococcus aureus* that now in its methicillin resistant form (MRSA) is arguably considered one of the most serious super bugs. Resistance to methicillin is most frequently obtained through the staphylococcal cassette chromosome (*scc*) containing the *mecA* penicillin binding protein gene [37]. This cassette is a large mobile genetic fragment that has even been determined to have been imported into a single MRSA clone on at least 23 occasions [37].

Vancomycin, which is primarily used to treat infections caused by gram-positive

bacteria that are unresponsive to other less toxic antibiotics [38], was first met with resistance in treating enterococci infections the late 1980s in Europe [39, 40]. A few years later, similar *vanA* vancomycin resistant enterococci (VRE) were detected in the Eastern United States, followed by cases world wide [41, 42]. The VanA-type of high-level resistance to vancomycin was the first to be described and is the most common [43]. Sequence analysis has determined that the *van* gene cluster is associated with a Tn*1546* transposon and is responsible for the dissemination to clinical *Enterococcus faecium* isolates [44]. The VanA-type of resistance has also been found in *S. aureus* isolates, even in MRSA isolates, where the resistance genes were also located on a plasmid-borne Tn*1546* [43].

3.5 Origins of Antibiotic Resistance

One hypothesis for the origin of antibiotic resistance genes is termed the "producer hypothesis". This hypothesis postulates that the natural producers of antibiotics also have a gene countering the effect of the antibiotics they produce [45]. Indeed, most of the successful, early antibiotics were isolated from strains in soil samples belonging to the genera *Streptomyces* [46], and many of the antibiotics used in the clinic today are structurally based on these compounds [47]. Consequently, these antibiotic-producing strains require immunity proteins in order for themselves to not be affected. One proposed role of these antibiotics is to help compete in soil for both space and food [46]. However, as only very few of the low molecular weight compounds secreted by microbes in the environment have identified antibiotic activity or even reach the concentrations required to act as an antibiotic, it is also speculated that these antibiotics are part of a larger class of signaling molecules [48]. In fact, some common clinically used antibiotics at sub-inhibitory concentrations even trigger the expression of virulence factors in the opportunistic pathogen *P. aeruginosa* [49]. In the antibiotics as warfare argument, other strains in close spatial contact with antibiotic producers would have a selective advantage over similar strains if they harbour the genes to counter-act the effects of these antibiotics. An advantage would exist whether their own antibiotic resistance genes evolved, or whether antibiotic resistance genes were acquired from the producer via horizontal gene transfer.

Antibiotic resistance also comes about through mutation, as mentioned earlier. On long evolutionary time-scales, this could include re-appropriation of a gene. However, during shorter courses of antibiotic treatment, this is primarily only relevant for

resistances that only require a small amount of mutations in a gene. It is horizontal gene transfer that gives prokaryotic genomes high plasticity, providing a faster route to antibiotic resistance [50, 51].

Interestingly, it is evident that antibiotic resistance is in fact an ancient phenomenon and predates our widespread use of antibiotics [52]. In a study by D'Costa and colleagues, two permafrost sediment cores taken from the Beringian forest in the Yukon territory of Canada, and estimated to be approximated 30,000 years old. DNA was extracted from this sample and interrogated for resistance genes via polymerase chain reaction (PCR). Resistance genes for several major classes of antibiotics were identified, and confirmed that ancient forms of the VanA vancomycin resistance protein had similar enzymatic function and structure as the modern form. The work highlights that antibiotic resistance is a naturally occurring phenomenon.

3.6 The Resistome

The collection of all genes conferring antibiotic resistance is called the "resistome" [53]. This includes all antibiotic resistance genes from pathogenic and commensal bacteria, from antibiotic producing bacteria, and cryptic genes that might not necessarily be expressed [53]. It additionally includes all precursor genes that could evolve from moderate into effective antibiotic resistance genes [53].

One of the first sampling of an environment's resistome was by D'Costa and colleagues of the soil resistome [54]. In this study they isolated spore-forming bacteria from the soil and screened this library against 21 different antibiotics, encompassing eight drug classes, including both natural and semisynthetic compounds. They found that every strain was in fact resistant to more than one antibiotic, and they observed resistances to every class of drugs tested.

In an alternative approach, Sommer and colleagues investigated the reservoir of antibiotic resistance genes in the human microbiota [55]. In addition to assessing the antibiotic resistance profiles of isolates cultured aerobically from the gut, they also employed a culture-independent strategy. Here, they isolated DNA from the human microbiota from saliva and fecal samples of healthy individuals. They sheared DNA into one- to three-kilobase (kb) fragments, cloned the DNA into an expression vector, and transformed the vector into an *E. coli* host strain. The resulting functional metagenomic libraries were screened via plating onto 13 different antibiotic-supplemented plates for metagenomic DNA fragments conferring antibiotic resistance. This culture-

independent approach led them to identify many resistance genes that had not been previously identified via sequencing.

Studies investigating the resistome have played a significant part in showing that a large pool of resistance genes exists. When sampling different environments, resistances can be found to antibiotics that they would never have been exposed to [54]. Given the possibility of mobilizing these genes via horizontal gene transfer, this is a worrying matter. In fact, this pool of resistance genes is consistent with the emergence of antibiotic resistances in clinical isolates occurring shortly after the introduction of an antibiotic [52].

Related the concept of the resistome, is the "subsistome". This termed was coined by J. Davies to describe the collection of genes that Dantas and colleagues identified which allow bacteria to degrade and subsist on antibiotics as their only carbon source [56, 9]. The ability for a human pathogen to subsist on antibiotics has been shown for a collection of *Salmonella* isolates [57]. Interestingly in many of the cases of subsistence, the identified gene was not related to a functional resistance gene.

3.7 Horizontal Gene Transfer in the Environment

Horizontal gene transfer is an ancient phenomenon, with transfer even occurring between the domains of life [58]. Much of bacterial innovation has been credited to horizontal gene transfer, whereas in eukaryotes, modification of existing genetic information plays a large role [50]. The first indication that horizontal gene transfer is significant for bacterial evolution was recognized in the mid 1950s, when genetically transferable antibiotic resistances were unexpectedly observed [9]. Yet prior to entering the full genomic sequencing age, early evolutionary biologist were still wary to cite horizontal gene transfer as the cause for inconsistencies in phylogenetic analyses [59]. Nevertheless, the significance of horizontal gene transfer was established as bacterial genome sequencing efforts intensified, yielding many indications of transfer [59].

Detecting horizontal gene transfer in the environment has been done in various ways. Studies have either been retrospective, based on identifying previous transfer events that have occurred, or prospective, based on direct observation of transfer events. Furthermore, various types of evidence exist, but are generally either sequence or phenotypic based. A short survey highlighting studies detecting horizontal gene transfer, or methods to detect horizontal gene transfer follows.

Shoemaker and colleagues conducted a retrospective study examining a collection

of *Bacteroides* isolates isolated over a period of more than 40 years to determine the extent of horizontal gene transfer within both clinical and community isolates [60]. First, they probed their isolates via DNA hybridization for tetracycline and erythromycin resistance genes. Apart from finding an increase in carriage of resistance genes over time, they found a high proportion of the resistant isolates carried only a small subset of the resistance genes they probed for. Next, they sequenced the identified resistance genes, which showed they had high sequence similarity. As the *Bacteroides* strains were only distantly related to each other, the high sequence similarity indicates that horizontal gene transfer was probable.

Other examples of retrospective studies include the work to determine the origin of the extended spectrum beta-lactamase CTX-M genes and their mobilization. Sequence analysis of the CTX-M gene clusters showed that the environmental source was from chromosomal beta-lactamases in the genus *Kluyvera* [61, 62, 63, 64, 65]. *In silico* analyses have shown that for some gene clusters they were even mobilized from their respective *Kluyvera* progenitors more than once [66]. The mobilization of these genes is largely attributed to IS elements, but also to a lesser extent to bacteriophages [66].

Studying horizontal gene transfer in a retrospective manner provides information on the interactions and evolutionary history of the genes, mobilizable elements and strains studied. With respect to studying the dissemination of antibiotic resistance, it highlights potential reservoirs of antibiotic resistance and illustrates possible dissemination routes to pathogens. However, challenges exist to certainly state the exact routes of dissemination retrospectively. Additionally, it is only possible to uncover successful events that have been conserved [67].

Perspective studies of horizontal gene transfer aim to observe transfer events. However, observing transfer events in natural settings can be difficult for various practical reasons. Observation not only requires that an event occurs, which can be infrequent, but it also requires proper sampling of the environment before and after transfer, amongst other factors. While controlled in vitro experiments have resulted in knowledge regarding rates of transfer and factors regarding transfer, the results may unfortunately not be transferable to natural environments, as the experimental conditions may not adequately reflect natural conditions. This has led to the development of artificial microcosms to study transfer, as well as seeding an environment with a particular donor strain and/or receiver strain.

When using a microcosm or adding specific donor bacteria to an environment,

there have traditionally been three approaches used for the detection of transfer: cultivation-based, fluorescence-based and molecular-based [67, 68]. A cultivation-based approach requires that the recipient bacteria are able to be cultivated, which is non-trivial in many environments. Furthermore, there must be a discernable phenotype associated with the mobilized element and the phenotype must be expressed in the recipient bacteria in order to detect the transfer [67]. Fluorescence-based approaches genetically modify mobilization elements to include expression of a fluorescent protein. This approach allows for spatial-temporal observation of transfer and has been used to identify transfer 'hot-spots' [67]. The last approach, a molecular-based approach, relies on quantitative PCR of distinct sequences to monitor an increase in mobilization element relative to the level of the donor bacterium [68].

Horizontal gene transfer has been observed in settings that did not use an artificial set-up with donor bacteria. Of the three previously mentioned approaches, it is primarily the culture-based approach that is amenable to studies where an environment is not seeded with a specific donor bacterium. Yet to observe a transfer event, thorough screening and characterization of the selected strains from the environment is required.

A new approach based on metagenomic data developed in the Banfield research group could avoid the issues and difficulties associated with perspective studies. This research group developed a method to assemble complete or near-complete genomes from a complex bacterial community [69]. Furthermore, using time-series sequencing data they were able to reliably assign mobilization elements to a particular species, but also to monitor their relative levels within the population. For example, they witnessed free phages in the gut environment, as well as phage integration and phage clearance. This method could provide an easier means to studying horizontal gene transfer in natural environment settings in the future.

3.8 Human Gut Microbiota

The human gut is home to one of the densest microbial ecosystems, namely the gut microbiota. In fact, most of the microbes belonging to the human microbiota are located within the intestinal tract, and they even out-number our own cells [70]. While the vast majority of microbes inhabiting the gut are from the bacterial domain, both Archaea and Eukarya are present as well [71]. Most studies give estimates of the bacterial species present in the human gut between 300 to 1000 species [71, 72, 73],

with some estimates of the total collective diversity assessed between 15,000 to 36,000 species [74]. The anaerobic environment of the gut is dominated by the Bacteroidetes and Firmicutes phyla, with the latter primarily containing members of the Clostridia class [72]. Also present, but found in much lower abundances are bacteria belonging to the Proteobacteria, Actinobacteria, Fusobacteria, and Verrucomicrobia phyla [72]. Facultative anaerobes, species that can perform both aerobic and anaerobic metabolism, constitute approximately 0.1% of the bacteria, consequently few Proteobacteria (including *E. coli*) would be expected [72].

Studies have sought to determine whether a core gut microbiota, a set of microbial species shared in common, exists. Qin et al. have indicated that some bacterial species are shared, but a subsequent study by the Human Microbiome Project were unable to identify a core [73, 75, 76]. However, the concept of a core microbiome, where genes are common in the gut microbiota, has proved to be more resilient, with genes encoding for similar metabolic functions shared across microbiomes [77, 76].

The gut microbiota has been envisioned as a microbial organ within a host organ, with many important functions for human health [78]. For example, the gut microbiota plays a role in the maturation of the adaptive immune system, protects the host against pathogens, synthesizes essential amino acids and vitamins, provides access to nutrients that are otherwise non-digestible, and influences energy storage [79, 71, 80, 78].

The role of the human gut microbiota has also been implicated in certain diseases. A study by Ley and colleagues focusing on obesity have found that there is a compositional difference of the major bacterial divisions, the Bacteriodetes and Firmicutes, with obese individuals having fewer Bacteriodetes and more Firmicutes than the lean control individuals [81]. Through diet intervention, the relative proportions of Bacteriodetes increased compared to the Firmicutes. In a different study by Turnbaugh and colleagues, they found that the obesity-associate gut microbiome has an increased potential for energy harvest, highlighting that the gut microbiome is an important component to consider, much like genetic factors and life-style [82].

Study of the gut microbiota has also been applied to bowel diseases, such as Crohn's disease and ulcerative colitis inflammatory bowel diseases. Frank and colleagues found that a subset of patients had a depletion of the normal commensal gut bacteria, particularly the Bacteroidetes and Firmicutes [74]. Gevers and colleagues sought to detect a possible disease signal within the microbiota in new-onset Crohn's disease, where the microbiota has not yet been perturbed due to treatment, and found

analysis of rectal-mucosal microbiota offers a potential early diagnosis [83].

The role the gut microbiota plays in the development of allergy has also been investigated. Studies have found that reduced diversity during the initial colonization of the gut after birth has been linked to the development of allergy later on [84, 85, 86].

3.9 Maturation of the Human Gut Microbiota

The infant intestinal tract is believed to be a bacteria-free environment before birth. The colonization process of the infant gut starts upon delivery, during its first exposures to bacteria, and converges towards a mature gut microbiota by age three [87]. Differences in the colonization and composition of the gut microbiota during this time has been linked to the mode of delivery [88, 89, 90], and to whether the infant was initially given breast milk or was formula fed [91, 88, 92].

The colonization process of the infant gut appears to undergo a series of shifts during the first year [93, 94]. In the Palmer *et al.* study of 14 infants, the shifts could not often be correlated to a specific age or event, apart from a shift towards an 'adult' microbiota upon the introduction of solid foods [93]. The Koenig *et al.* study had only one infant, albeit with more sampling points, and were largely able to attribute shifts to changes in diet, illness or antibiotic treatment [94]. This result does however corroborate with the set of fraternal twins in the Palmer et al. study, as they displayed a highly similar temporal patterns, indicating that the shifts were in response to an environmental exposure.

Comparing the composition of the infant gut microbiota revealed that there was greater variation amongst them, than was present in the adult gut microbiota [87]. The infant gut microbiota was also surprisingly found to not be more similar to their own parents' gut microbiota, than to any other parent's gut microbiota [93]. While these results may indicate a disorderly infant gut microbiota, colonization does follow a trend from the initial aerobes, such as *Staphylococcus*, *Streptococcus* and Enterobacteria, to the strict anaerobes [93]. Many studies have found that *Bifidobacterium* tend to be a dominant genus in the infant gut microbiota, apart from the Palmer *et al.* study [87, 95, 96, 93]. In the Palmer *et al.* study, *Bacteroides* dominated early in some infants, yet was present for the majority of infants by the end of the first year [93].

3.10 Studying the Human Gut Microbiota

The human gut microbiota has been studied for over a hundred years [97], with renewed interest upon advances in technology. The earliest studies were cultivation-based. For example, Tissier and colleagues were the first to describe *Bifidobacterium* when they cultured infant fecal samples, and noted that these "bifidus" bacteria were dominant and more numerous from infants fed breast milk rather than formula [97, 98]. Research continued in isolating and describing more species from the gut, yet it was first in the 1930s when it was acknowledged that majority of the viable bacteria in the gut were anaerobic [99].

The next major advances in studying the gut microbiota came with the emergence of molecular methods and sequencing. These methods provide culture-independent approaches to study the composition and diversity of the gut microbiota, without the concern of neglecting the fraction that are not amenable to cultivation.

The initial sequencing studies targeted a specific gene or set of genes. The target was primarily the 16S rRNA gene due to its ubiquitous distribution and the phylogenetic information it provides [100], but other genes with a more limited distribution have been targeted to screen for certain functionalities [101]. It may also not even be the entire gene that is targeted. Some studies have targeted one or more hypervariable regions within the 16S rRNA gene to assess the diversity of their samples, yet still rely on databases of full-length sequences for taxonomic assessment [102]. This short sequence tag approach made efficient use of the high-throughput, yet relatively short reads that were produced by next generation sequencing machines. However, as expected, full 16S rRNA gene sequencing provides higher phylogenetic resolution. It is important to note that there is often a PCR amplification bias which can result from differing amplification efficiencies of templates, self-annealing of the most abundant transcripts in the late amplification steps, or from non-linear amplification [103, 104, 105].

Another group of methods for assessing the bacterial community composition and diversity is based on hybridization of oligonucleotides. Similar to the targeted amplicon approach, oligonucleotides are designed to target a specific region of a gene. However, here taxonomic classification is via the detection of hybridization, and not via sequence analysis. In fluorescent *in situ* hybridization (FISH), oligonucleotide probes are designed to identify certain taxa and are fluorescently labeled. These probes are added to bacterial samples to hybridize to the DNA of the bacteria, and then the

bacteria can then be sorted using flow cytometry [106]. FISH studies can provide spatial resolution by first fixating a sample prior to hybridization, and subsequently visualizing the fluorescent probes [107]. DNA microarrays such as the PhyloChip have been developed for taxonomic classification based on hybridization of amplified 16S rRNA genes [108]. These hybridization methods can only detect taxa that have been targeted by the affixed oligonucleotides. FISH can typically be run for the detection of 6-8 genus or family level groups [100]. DNA microarrays can detect many more taxa simultaneously, yet quantification of DNA microarrays can be an issue [100].

Lastly, 'finger print' types of methods are an inexpensive and relatively simple way to assess the dominant groups in a microbial community. These methods include terminal restriction fragment length polymorphism (T-RFLP), denaturing and temperature gradient gel electrophoresis (DGGE and TGGE). In short, these methods also target a gene, also typically the 16S rRNA gene, followed by subsequent separation of the variants via gel electrophoresis [102]. Unfortunately, the dynamic range is not as large as the other methods, and results are often not comparable between studies [102].

As sequencing technology has improved to generate more reads of longer lengths, and the cost of sequencing has decreased, the focus has shifted to full metagenomic studies. In contrast to targeted amplicon sequencing where one or a few specific genes are sequenced, all DNA from the microbial community is extracted and sequenced. The data can be assembled into a metagenome, a type of environmental genome. Methods to characterize the metagenome based on composition have been developed [109, 110, 111, 112], yet challenges remain to correctly ascertain which bacteria contained which DNA. However, the leading benefit of metagenomic sequencing comes from analyzing the function of the genetic content to determine the full capacity of the metagenome [102]. This approach follows the notion that gene-based core metagenome exists, and it is the function of the metagenome as a whole that is interesting to study. However, assigning gene function can be challenging due to a lack of close matches in reference databases [73, 101]. On the other hand, metagenomic reads need not always be assembled, with function subsequently assigned to identified open reading frames. Alternatively, reads can be mapped to genes or genomes to determine abundance or absence in sample sets [73, 113, 101].

While many advances have been made to improve our understanding of the gut microbiota, in both healthy and diseased states, it is still necessary to test the hypotheses generated in model systems. To this end, there is renewed interest and need

for model systems and for further development of culture-based methods, including improving cultivability of the human gut microbiota.

Germ-free and gnotobiotic mice have been employed as models to study the human gut microbiota and its relationship to the host. Germ-free mice are born in asceptic conditions and reared in special facilities where all air, food and water is sterilized. Gnotobiotic mice are defined as those where the natural microbiota has been replaced by a defined set of bacteria. In practice, gnotobiotic mice are attained by introducing a specific set of bacteria to a germ-free mouse, often through gavaging. The use of these animals allows for comparisons between germ-free, gnotobiotic and normal gut microbiota mice. For example, one study found that germ-free mice were less anxious and more active than the mice with a normal gut microbiota [114]. Other studies analyzing facultative anaerobes have used mice with a normal gut microbiota that is treated with streptomycin. Streptomycin-treatment provides an opportunity for the facultative bacteria of interest, marked with streptomycin resistance, to colonize the gut while the anaerobic bacteria remain largely intact [115, 116].

Cultivating bacteria isolated from the gut can not only be labourious, but it can be very technically challenging to establish the correct culture conditions. In fact, this has led modern sequence-based studies to proclaim that a large fraction of the human gut may be recalcitrant to culture [72, 117, 118, 119, 120]. This assertion could be due to a variety of reasons, including the difficulties in culturing, such as providing the correct nutrients and physiological conditions, or lacking symbionts when attempting to grow in pure cultures [77]. Additionally, disparities in abundances can leave less abundant species not sampled, while disparities in growth could leave slower growing organisms less likely to be isolated. Even with these pessimistic views, there have been a few recent advances in culturing techniques. The new techniques include the encapsulation of bacteria into microdroplets [121, 122], providing the natural or closely simulated environment by growing the microorganism in a diffusion chamber allowing for exchange with the environment [123, 124], more effective culture media, including the use of other solidifying agents instead of agar [125, 126, 127]. Even the concept of culturomics has been introduced as a high-throughput method combining new culturing methods with mass spectrometry for detection of newly cultivated species from the gut microbiota [128]

Even with difficulties in culturing bacteria from the human gut microbiota, isolating and culturing gut microbes is an important step towards furthering our understanding of the human gut microbiota. Through cultivation-based studies, we can

assess interactions between bacterial species [129, 130]. Employing full genome sequencing on these isolates is less challenging than attempting to discern full genomes from metagenomic data and this will in turn enhance reference databases. This can assist in metagenomic studies, or can be mined for new functionalities or to identify new therapeutic targets [101].

3.11 Lineage Dynamics in the Human Gut Microbiotia

Investigating the human gut microbiota with strain-level of resolution provides insight into the individual characteristics of the strains and the roles they play in the gut microbiota. For example, some species present in the gut microbiota may also include strains that are virulent, such as *E. coli*. Similarly, only certain strains of *Bifidobacterium* can protect the host from enterohaemorrhagic *E. coli* infection [131]. Fukuda *et al.* found that these protective *Bifidobacterium* strains contain an ATP-binding cassette-type of carbohydrate transporter, which contributes to acetate production by these strains in the distal colon where glucose has been exhausted [131].

The ability to study the human gut microbiota at the strain-level is also important to study the population dynamics of this environment. Some studies have taken to examining lineage turn over in the gut to understand the colonization stability. In the 1950s, Sears and colleagues conducted longitudinal studies examining *E. coli* strain composition and persistence in the adult gut, with strains classified based on their O-antigen group [132, 133]. Here they termed resident and transient strains as strains that persist in the gut over long periods of time and strains that maintain a tenure of only a few days or weeks, respectively [132].

In similar studies, Wold, Adlerberth and colleagues investigated strain turn over during the colonization of the infant gut for *E. coli* [134, 135, 136, 137], and for *S. aureus* [138]. This work is part of the ALLERGYFLORA study where they enrolled Swedish infants to study the link between colonization of gut and the onset of allergies later in life. They quantitatively cultured all major groups of facultative and anaerobic bacteria from fecal samples and identified strains via random amplified polymorphic DNA (RAPD) typing. In their study with 70 infants, they discovered that there were 2.1 *E. coli* strains sampled per infant for the year [134], and that the B2 phylogenetic group had superior capacity to persist in the gut [135]. In the studies regarding *S. aureus*, they found that certain combinations of virulence factors, such as the *agr* allele I and the *spa* clonal cluster 630 (spa-CC 630), promoted gut colonization as

these factors were found in all resident strains [138].

Banfield and colleagues have developed a culture independent approach using time-series metagenomic samples to gain strain-level resolution. In a sequence of papers, they employed this approach to investigate the colonization of the infant gut [139, 69, 140]. In general, their approach is based on binning together groups of genomes with similar abundances and to re-assemble these sequences together, separate from the rest [69]. The time series abundance profiles provide additional information for grouping contigs correctly into genomes or similar strains. Using this method, they were able to assemble two near-complete genomes and nine partially complete genomes for bacteria that comprised as little as 1% and 0.05% of the microbial community from one infant[69]. They also tracked the relative abundances of three *Staphylococcus epidermis* strains and phages distinctly infecting each of them, and related a decline in one phage to the increase in abundance of the *S. epidermis* it infected and noted this strain was the only one that contained an abortive infection bacteriophage system [69]. In another study, they assembled two distinct *Citrobacter* strains and linked strain-level differences to the dominance of one strain to possible metabolic and host interaction traits via comparative genomics [139].

Their method to assemble complete- or near-complete genomes from time-series metagenomic sampled allows for insights into individual strain contributions of the gut microbiota, while tracking strain abundances over time. Although strain-level resolution is achieved via their method, it is possible to fail to identify some events that change the abundance of sequences, such as prophage integration and curing, acquisition or loss of plasmids, and genomic insertions or deletions. As the method relies on sequence abundances to group together strain specific sequences, it is possible that deficiencies in sampling could result in the inability to connect changes in abundances to a particular strain. With sufficient sampling, it could be possible to detect such events, but great care in interpreting the time-series abundance profiles must be taken to ensure that these events are not missed.

3.12 Horizontal Gene Exchange within the Human Gut Microbiota

The human gut is regarded as a suitable environment for gene exchange, and is even considered to serve as a hub. The focus on the role the bacteria in the intestinal tract play in the dissemination of antibiotic resistance gene is due to many factors. The

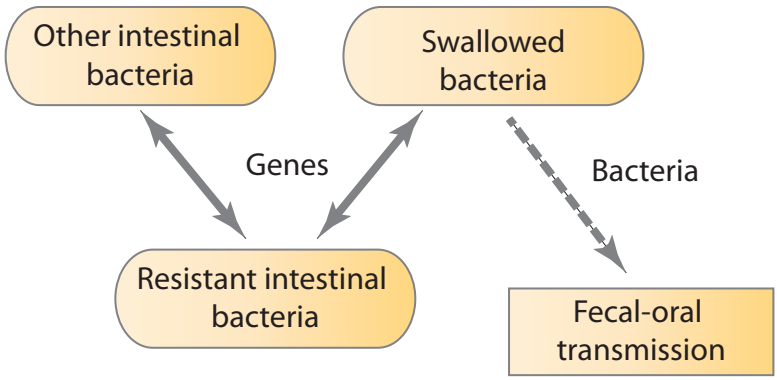


Figure 3.2 | The resistance gene reservoir hypothesis and the role of the human gut microbiota. Figure adapted from Salyers *et al.*, 2004, Trends Microbiol.

first factor is that the gut microbiota has been identified as a reservoir of antibiotic resistance genes [55]. A second factor critical to the dissemination of these resistance genes is that the human gut has the high cell density required to facilitate horizontal gene transfer via conjugation [12]. And lastly, it may not only be the resident intestinal bacteria that is partaking in horizontal gene transfer, but also other bacteria that are merely transiting through the gut that can be involved in gene exchange [141] (Fig. 3.2). As many of the commensal bacteria in the gut are opportunistic pathogens, there is concern that these bacteria may receive virulence factors or antibiotic resistance genes, making them more pathogenic or difficult to treat [141]. Additionally, concerns exist regarding the potential of non-colonizing bacteria passing through the gut may disseminate or acquire genes before returning to their normal body sites [141].

The possibility of the gut serving as an environment for the dissemination of antibiotic resistance genes was first investigated in the late 1960s and 1970s. Quite a few studies employed experimental setups to assess whether *in vivo* transfer of "R factors" (resistance factors) was possible in the gut [142, 143]. In the studies by Smith

et al. and Anderson *et al.*, volunteers were orally administered genetically marked E. coli receiver strains, followed by prospective *E. coli* donator strains. Fecal samples collected from the volunteers were cultured to observe if the resistance phenotype of the receiver strain changed. In the Smith *et al.* study, they used *E. coli* strains isolated from humans and from a variety of animals. While they observed transfer from both human and animal strains in their one test subject, the animal strains of *E. coli* were poorer at colonizing the gut and had to be given to at much higher numbers than humans would be expected to consume naturally. In the Anderson *et al.* study, they were unable to detect transfer of resistance genes in any of their volunteers without antibiotic treatment. After a course of ampicillin or tetracycline, transfer of resistance genes to their tagged receiver strain was observed in three out of the five volunteers.

Other early studies used epidemiological data to infer that *in vivo* transfer of resistance genes likely occurred. In a case by Ferrar and colleagues, a probable instance of interbacterial transfer was recognized during an outbreak of shigellosis in a nursery caused by *Shigella sonnei* [144]. The pattern of antibiotic susceptibility of the isolated *S. sonnei* was initially the same for all patients; all isolates were resistant streptomycin, tetracycline, and ampicillin, but sensitive to kanamycin. However, after therapy with kanamycin, the antibiotic susceptibility profile of *S. sonnei* changed in one infant to include kanamycin resistance. Both kanamycin sensitive and resistant *S. sonnei* isolates were capable of transmitting antibiotic resistance to *E. coli*. The pattern of resistance in the *E. coli* progeny indicated that there were two distinct R factors: one carrying streptomycin, tetracycline and ampicillin resistance, and the other kanamycin resistance. The observation that the determinant for kanamycin resistance was on a separate mobilisable element provided evidence that the *S. sonnei* strain acquired kanamycin resistance via horizontal gene transfer in the gut.

Petrocheilou and colleagues reported a possible transfer of a plasmid containing antibiotic resistance genes between co-existing *E. coli* O-antigen groups in study published in 1976 [145]. The main distinction in this study is that their evidence for transfer is not based on an experimental setup where volunteers were fed marked strains, nor does it come from epidemiological data, and lastly it is not in conjunction with antibiotic treatment. However, the proposed transconjugant was only observed after the suspected transfer occurred. Evidence that a plasmid carrying antibiotic resistance genes transferred between the different *E. coli* O-antigen groups is based on molecular methods of the time, namely *Eco*RI restriction digests separated on gel,

and phenotypic similarities such as production of F-pili, *fi* character, similar partial plasmid incompatibility and resistance phenotype [145]. Still, as this strain with a similar plasmid was only observed after the suspected transfer occurred, it is not conclusive whether the transfer occurred *in situ* of the intestinal tract or if the strain already harboured such a plasmid.

These early studies mentioned relied to a large extent on selective culturing using antibiotic resistance phenotypes to detect transfer. This can be problematic when determining if a particular strain has indeed changed its resistance profile due to acquisition of antibiotic resistance genes. For the Anderson *et al.* study, they chromosomally tagged the hosts' own *E. coli* strains, and evidence for the transfer of antibiotic resistance genes was only based on a positive culture of an isolate with a resistance phenotype that was a union of both the marker and the donator. Although the receiver strain was isolated from each of the volunteers, there was no mention of the diversity of *E. coli* lineages, nor any other phenotypic controls for detecting transfer. In the Smith *et al.* study, identification of the receiving strain was based on a chromosomal antibiotic resistance marker, haemolysis, and slide-agglutination tests with an antiserum prepared against it. In the study by Petrocheilou and colleagues, they gained a slightly higher level of resolution on the *E. coli* strain identities of the microbiota, yet the extent of lineage identification was based on O-antigen groups, some of which varied in their antibiotic resistance profile over the sampling time. Farrar and co-authors stated that as they were unable to colicin type the *S. sonnei* isolates they are likely the same strain. Yet even a study from that time found that around 40% of the *S. sonnei* strains were not possible to colicin type [146].

Additionally, for the reports where transfer occurred in an unperturbed gut, where volunteers or patients hadn't ingested a donor strain, even discerning whether the resistance genes were located on the same plasmid proved to be challenging. Petrocheilou and colleagues performed transduction experiments to conclude that the resistance markers were cotransducible and thus on the same plasmid [145]. Alternatively, Farrar and colleagues examined the pattern of the resistance profiles of the transconjugant progeny to discern that there were two plasmids [144].

Current studies have enjoyed the benefit of more advanced methods affording a higher resolution. For example, in a study by Trobos and colleagues to investigate whether the sulphonamide resistance gene *sul2* could be transferred between *E. coli* in the gut, strains were identified by pulsed-field gradient gel electrophoresis (PFGE); a method of gel electrophoresis that separates long DNA molecules by alternating

electrical pulses [147, 148]. Although this study intended on observing transfer between ingested donator and receiver strains, only one case of transfer amongst the six volunteers was observed. It was infact a native sulphonamide resistant *E. coli* strain transferring a plasmid to their receiver strain. Additionally, the sequencing of *sul2* showed that it was identical to that of the native *E. coli*'s own. The ability to 'fingerprint' the strains, combined with sequencing of the resistance determinant provided compelling evidence for the transfer.

In the Lester *et al.* study, they examined whether transfer of the *vanA* vancomycin resistance gene can occur in the human intestine from animal to human origin isolates of the gram-positive *Enterococcus faecium*, in the absence of selective pressure [149]. Transconjugants were observed in three out of the six volunteers, and presence of the *vanA* gene was confirmed via PCR. Additionally, a point mutation in the *vanX* gene was also confirmed in the donor and transconjugants. PFGE fingerprints of the strains showed that the resistant *in vivo* transconjugants had an additional band compared to non-resistant strain, and that both the *in vivo* and *in vitro* transconjugants had similar profiles.

While both of these studies used artificial experimental setups to determine if transfer of resistance genes could occur *in situ* of the gut, they applied more modern methods to verify and confirm the strains and resistance determinants involved in transfer. With respect to plasmid identification, sequencing segments or even an entire plasmid has become commonplace. Furthermore, PCR-based assays for determining the incompatability types of plasmids have been developed based on sequence information and can be used for plasmid group identification [150].

Very few recent reports exist documenting possible transfer *in situ* of the human gut, in non-artificial setups. One such report by Bidet and colleagues, documented the transfer of a plasmid encoded AAC-1 beta-lactamase gene from a *Klebsiella pneumonia* strain to an *E. coli* strain in the gut of an infant [151]. A fecal sample was analysed before the infant underwent surgery, yielding a cefotaxime-resistant *K. pneumonia* strain. The infant later developed a urinary tract infection due to an ampicillin-resistant *E. coli* at day 30 and 45 post surgery, and was treated with cefotaxime and gentamicin both times. At day 45, another fecal sample was analyzed yielding cefotaxime-resistant resistant *K. pneumonia* and *E. coli* strains. The beta-lactamase gene variants were screened and sequenced, revealing TEM-1 in all isolates, and AAC-1b in the cefotaxime-resistant *K. pneumonia* and *E. coli* isolates. Strain identities were verified via PFGE analysis, and plasmid identity was confirmed via restriction

digest of plasmid DNA separated by gel electrophoresis.

A similar report documenting transfer of resistance genes in the infant gut while undergoing antibiotic treatment is from Karami and colleagues [152]. The infant in this report was enrolled in the ALLERGYFLORA study, which set to examine whether a link exists between the initial colonization of the gut to the onset of allergies later on in life. Two distinct *E. coli* lineages colonized the gut within the first month of life; one was resistant to ampicillin, and the other one was sensitive to ampicillin. The *E. coli* lineage identities were confirmed via random amplified polymorphic DNA (RAPD) typing. The infant developed a urinary tract infection within the first month, and was treated first with trimethoprim, then ampicillin, and finally amoxicillin. After antibiotic treatment with ampicillin and amoxicillin, both *E. coli* lineages were resistant to ampicillin. The beta-lactamase variants were detected via isoelectric focusing, and a TEM-1 type beta-lactamase that conferred resistance was confirmed via sequencing. Plasmid identification was done through restriction fragment length polymorphism (RFLP) analysis, revealing a plasmid of approximately 40kb, and presence of the TEM-1 type beta-lactamase was confirmed via Southern-blot hybridization.

These reports by Bidet *et al.* and Karami *et al.* demonstrate that transfer of resistance genes in the gut is possible in a non-artificial experimental set-up. However in both of the observations, transfer was observed while the infants were undergoing antibiotic treatment. The other studies by Trobos *et al.* and Lester *et al.*, demonstrate that transfer is possible in adult gut without the need for antibiotic pressure. However, these studies employed artificial experimental set-ups where volunteers ingested high titres of potential donor and/or receiver strains. These conditions do not accurately represent the natural state of the gut microbiota, and similarly the amounts of ingested bacteria do not likely correspond to the amount of bacteria ingested on food [142]. The question whether transfer occurs in the unperturbed gut without antibiotic still remains to be answered. Furthermore, follow up analysis to investigate the fate of the horizontally acquired resistant bacteria in the gut can bring insight into whether it can remain colonized in the gut, and if any adaptations have occurred. This potential for the reservoir of resistance genes in the gut to disseminate and expand within the gut, and for the horizontally acquired resistant bacteria to survive is crucial in the risk assessment of the gut as a hub of antibiotic resistance genes exchange.

3.13 References

[1] World Health Organization. Antimicrobial resistance: global report on surveillance. 2014.

[2] L L Silver. Challenges of Antibacterial Discovery. *Clinical microbiology reviews*, 24(1):71–109, January 2011.

[3] H Nikaido. Molecular Basis of Bacterial Outer Membrane Permeability Revisited. *Microbiology and Molecular Biology Reviews*, 67(4):593–656, December 2003.

[4] Ayush Kumar and Herbert P Schweizer. Bacterial resistance to antibiotics: active efflux and reduced uptake. *Advanced drug delivery reviews*, 57(10):1486–1513, July 2005.

[5] M Oram and L M Fisher. 4-Quinolone resistance mutations in the DNA gyrase of Escherichia coli clinical isolates identified by using the polymerase chain reaction. *Antimicrobial agents and chemotherapy*, 35(2):387–389, February 1991.

[6] F M Barnard and A Maxwell. Interaction between DNA Gyrase and Quinolones: Effects of Alanine Mutations at GyrA Subunit Residues Ser83 and Asp87. *Antimicrobial agents and chemotherapy*, 45(7):1994–2000, July 2001.

[7] B Weisblum. Erythromycin resistance by ribosome modification. *Antimicrobial agents and chemotherapy*, 39(3):577–585, March 1995.

[8] J Davies. Inactivation of antibiotics and the dissemination of resistance genes. *Science (New York, N.Y.)*, 264(5157):375–382, April 1994.

[9] J Davies and D Davies. Origins and Evolution of Antibiotic Resistance. *Microbiology and Molecular Biology Reviews*, 74(3):417–433, August 2010.

[10] Shalini Mathur and Rameshwar Singh. Antibiotic resistance in food lactic acid bacteria—a review. *International Journal of Food Microbiology*, 105(3):281–295, December 2005.

[11] P Courvalin. Transfer of antibiotic resistance genes between gram-positive and gram-negative bacteria. *Antimicrobial agents and chemotherapy*, 38(7):1447–1451, July 1994.

[12] Jennifer R Huddleston. Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. *Infection and Drug Resistance*, page 167, June 2014.

[13] Carlos Canchaya, Ghislain Fournous, Sandra Chibani-Chennoufi, Marie-Lise Dillmann, and Harald Brüssow. Phage as agents of lateral gene transfer. *Current Opinion in Microbiology*, 6(4):417–424, August 2003.

[14] Rémi Fronzes, Peter J Christie, and Gabriel Waksman. The structural biology of type IV secretion systems. pages 1–12, October 2009.

[15] A Babic, A B Lindner, M Vulic, E J Stewart, and M Radman. Direct Visualization of Horizontal Gene Transfer. *Science (New York, N.Y.)*, 319(5869):1533–1536, March 2008.

[16] An-Chi Shu, Chien-Chen Wu, Yi-Yang Chen, Hwei-Ling Peng, Hwan-You Chang, and Tri-Rung Yew. Evidence of DNA Transfer through F-pilus Channels during Escherichia coliConjugation. *Langmuir*, 24(13):6796–6802, July 2008.

[17] I Chen. The Ins and Outs of DNA Transfer in Bacteria. *Science (New York, N.Y.)*, 310(5753):1456–1460, December 2005.

[18] Maria Smorawinska, Magdalena Szuplewska, Piotr Zaleski, Paweł Wawrzyniak, Anna Maj, Andrzej Plucienniczak, and Dariusz Bartosik. Mobilizable narrow host range plasmids as natural suicide vectors enabling horizontal gene transfer among distantly related bacterial species. *FEMS Microbiology Letters*, 326(1):76–82, November 2011.

[19] Christopher M Thomas and Kaare M Nielsen. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nature Reviews Microbiology*, 3(9):711–721, September 2005.

[20] Patrick Seitz and Melanie Blokesch. Cues and regulatory pathways involved in natural competence and transformation in pathogenic and environmental Gram-negative bacteria. *FEMS Microbiology Reviews*, 37(3):336–363, September 2012.

[21] Julie A Perry and Gerard D Wright. The antibiotic resistance "mobilome": searching for the link between environment and clinic. *Frontiers in microbiology*, 4, 2013.

[22] M Kohiyama. Bacterial Sex: Playing Voyeurs 50 Years Later. *Science (New York, N.Y.)*, 301(5634):802–803, August 2003.

[23] G G Perron, A E G Lee, Y WANG, W E Huang, and T G Barraclough. Bacterial recombination promotes the evolution of multi-drug-resistance in functionally diverse populations. *Proceedings of the Royal Society B: Biological Sciences*, 279(1733):1477–1484, March 2012.

[24] S H Gillespie. Evolution of Drug Resistance in Mycobacterium tuberculosis: Clinical and Molecular Perspective. *Antimicrobial agents and chemotherapy*, 46(2):267–274, February 2002.

[25] Lei Yang, Lars Jelsbak, Rasmus Lykke Marvig, Søren Damkiær, Christopher T Workman, Martin Holm Rau, Susse Kirkelund Hansen, Anders Folkesson, Helle Krogh Johansen, Oana Ciofu, Niels Høiby, Morten O A Sommer, and Søren Molin. Evolutionary dynamics of bacteria in a human host environment. *Proceedings of the National Academy of Sciences*, 108(18):7481–7486, May 2011.

[26] Michael D Iseman. Evolution of drug-resistant tuberculosis: a tale of two species. *Proceedings of the National Academy of Sciences of the United States of America*, 91(7):2428–2429, 1994.

[27] World Health Organization. Anti-tuberculosis drug resistance in the world: fourth global report. 2008.

[28] J B Lyczak, C L Cannon, and G B Pier. Lung Infections Associated with Cystic Fibrosis. *Clinical microbiology reviews*, 15(2):194–222, April 2002.

[29] C Koch and N HØIBY. Diagnosis and treatment of cystic fibrosis. *Respiration; international review of thoracic diseases*, 67(3):239–247, 2000.

[30] Martin H Rau, Rasmus Lykke Marvig, Garth D Ehrlich, Søren Molin, and Lars Jelsbak. Deletion and acquisition of genomic content during early stage adaptation of Pseudomonas aeruginosa to a human host environment. *Environmental Microbiology*, 14(8):2200–2211, June 2012.

[31] Rasmus Lykke Marvig, Helle Krogh Johansen, Søren Molin, and Lars Jelsbak. Genome Analysis of a Transmissible Lineage of Pseudomonas aeruginosa Reveals Pathoadaptive Mutations and Distinct Evolutionary Paths of Hypermutators. *PLoS Genetics*, 9(9):e1003741, September 2013.

[32] O Ciofu, B Giwercman, N HØIBY, and S S Pedersen. Development of antibiotic resistance in Pseudomonas aeruginosa during two decades of antipseudomonal treatment at the Danish CF Center. *APMIS*, 102(7-12):674–680, 1994.

[33] P M Hawkey and A M Jones. The changing epidemiology of resistance. *Journal of Antimicrobial Chemotherapy*, 64(Supplement 1):i3–i10, August 2009.

[34] F Perez, A M Hujer, K M Hujer, B K Decker, P N Rather, and R A Bonomo. Global Challenge of Multidrug-Resistant Acinetobacter baumannii. *Antimicrobial agents and chemotherapy*, 51(10):3471–3484, September 2007.

[35] A Y Peleg, H Seifert, and D L Paterson. Acinetobacter baumannii: Emergence of a Successful Pathogen. *Clinical microbiology reviews*, 21(3):538–582, July 2008.

[36] M G Smith, T A Gianoulis, S Pukatzki, J J Mekalanos, L N Ornston, M Gerstein, and M Snyder. New insights into Acinetobacter baumannii pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis. *Genes & Development*, 21(5):601–614, March 2007.

[37] Ulrich Nübel, Philippe Roumagnac, Mirjam Feldkamp, Jae-Hoon Song, Kwan Soo Ko, Yhu-Chering Huang, Geoffrey Coombs, Margaret Ip, Henrik Westh, and Robert Skov. Frequent emergence and limited geographic dispersal of methicillin-resistant Staphylococcus aureus. *Proceedings of the National Academy of Sciences of the United States of America*, 105(37):14130–14135, 2008.

[38] Robert C Moellering. Vancomycin: a 50-year reassessment. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 42 Suppl 1:S3–4, January 2006.

[39] A H Uttley, C H Collins, J Naidoo, and R C George. Vancomycin-resistant enterococci. *Lancet*, 1(8575-6):57–58, January 1988.

[40] Roland Leclercq, Eliane Derlot, Jean Duval, and Patrice Courvalin. Plasmid-Mediated Resistance to Vancomycin and Teicoplanin in Enterococcus Faecium. *The New England journal of medicine*, 319(3):157–161, July 1988.

[41] T R Frieden, S S Munsiff, D E Low, B M Willey, G Williams, Y Faur, W Eisner, S Warren, and B Kreiswirth. Emergence of vancomycin-resistant enterococci in New York City. *Lancet*, 342(8863):76–79, July 1993.

[42] J M Bell, J C Paton, and J Turnidge. Emergence of vancomycin-resistant enterococci in Australia: phenotypic and genotypic characteristics of isolates. *Journal of clinical microbiology*, 36(8):2187–2190, August 1998.

[43] B Perichon and P Courvalin. VanA-Type Vancomycin-Resistant Staphylococcus aureus. *Antimicrobial agents and chemotherapy*, 53(11):4580–4587, October 2009.

[44] M Arthur, C Molinas, F Depardieu, and P Courvalin. Characterization of Tn1546, a Tn3-related transposon conferring glycopeptide resistance by synthesis of depsipeptide peptidoglycan precursors in Enterococcus faecium BM4147. *Journal of bacteriology*, 175(1):117–127, January 1993.

[45] Raoul Benveniste and Julian Davies. Aminoglycoside antibiotic-inactivating enzymes in actinomycetes similar to those present in clinical isolates of antibiotic-resistant bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 70(8):2276–2280, 1973.

[46] David A Hopwood. Streptomyces in nature and medicine: the antibiotic makers. 2007.

[47] Gautam Dantas and Morten OA Sommer. Ecological and Clinical Consequences of Antibiotic Subsistence by Environmental Microbes. *Antimicrobial Resistance in the Environment*, page 29, 2011.

[48] G Yim, H Huimi Wang, and J Davies FRS. Antibiotics as signalling molecules. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1483):1195–1200, July 2007.

[49] J F Linares, I Gustafsson, F Baquero, and J L Martinez. Antibiotics as intermicrobial signaling agents instead of weapons. *Proceedings of the National Academy of Sciences of the United States of America*, 103(51):19484–19489, 2006.

[50] H Ochman, J G Lawrence, and E A Groisman. Lateral gene transfer and the nature of bacterial innovation. *Nature*, 405(6784):299–304, May 2000.

[51] Eva M Top and Dirk Springael. The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Current opinion in biotechnology*, 14(3):262–269, June 2003.

[52] Vanessa M D'Costa, Christine E King, Lindsay Kalan, Mariya Morar, Wilson W L Sung, Carsten Schwarz, Duane Froese, Grant Zazula, Fabrice Calmels, Regis Debruyne, G Brian Golding, Hendrik N Poinar, and Gerard D Wright. Antibiotic resistance is ancient. *Nature*, 477(7365):457–461, April 2012.

[53] Gerard D Wright. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature Reviews Microbiology*, 5(3):175–186, March 2007.

[54] Vanessa M D'Costa, Katherine M McGrann, Donald W Hughes, and Gerard D Wright. Sampling the antibiotic resistome. *Science (New York, N.Y.)*, 311(5759):374–377, January 2006.

[55] M O A Sommer, G Dantas, and G M Church. Functional Characterization of the Antibiotic Resistance Reservoir in the Human Microflora. *Science (New York, N.Y.)*, 325(5944):1128–1131, August 2009.

[56] Gautam Dantas, Morten OA Sommer, Rantimi D Oluwasegun, and George M Church. Bacteria subsisting on antibiotics. *Science (New York, N.Y.)*, 320(5872):100–103, 2008.

[57] A E Barnhill, K E Weeks, N Xiong, T A Day, and S A Carlson. Identification of Multiresistant Salmonella Isolates Capable of Subsisting on Antibiotics. *Applied and Environmental Microbiology*, 76(8):2678–2680, April 2010.

[58] James R Brown. Ancient horizontal gene transfer. *Nature Reviews Genetics*, 4(2):121–132, February 2003.

[59] J Peter Gogarten, W Ford Doolittle, and Jeffrey G Lawrence. Prokaryotic evolution in light of gene transfer. *Molecular Biology and Evolution*, 19(12):2226–2238, 2002.

[60] N B Shoemaker, H Vlamakis, K Hayes, and A A Salyers. Evidence for Extensive Resistance Gene Transfer among Bacteroides spp. and among Bacteroides and Other Genera in the Human Colon. *Applied and Environmental Microbiology*, 67(2):561–568, February 2001.

[61] J W Decousser, L Poirel, and P Nordmann. Characterization of a Chromosomally Encoded Extended-Spectrum Class A beta -Lactamase from Kluyvera cryocrescens. *Antimicrobial agents and chemotherapy*, 45(12):3595–3598, December 2001.

[62] C Humeniuk, G Arlet, V Gautier, P Grimont, R Labia, and A Philippon. β-Lactamases of Kluyvera ascorbata, Probable Progenitors of Some Plasmid-Encoded CTX-M Types. *Antimicrobial agents and chemotherapy*, 46(9):3045–3049, September 2002.

[63] A B Olson, M Silverman, D A Boyd, A McGeer, B M Willey, V Pong-Porter, N Daneman, and M R Mulvey. Identification of a Progenitor of the CTX-M-9 Group of Extended-Spectrum -Lactamases from Kluyvera georgiana Isolated in Guyana. *Antimicrobial agents and chemotherapy*, 49(5):2112–2115, April 2005.

[64] M M Rodriguez, P Power, H Sader, M Galleni, and G Gutkind. Novel Chromosome-Encoded CTX-M-78 -Lactamase from a Kluyvera georgiana Clinical Isolate as a Putative Origin of CTX-M-25 Subgroup. *Antimicrobial agents and chemotherapy*, 54(7):3070–3071, June 2010.

[65] L Poirel, P Kampfer, and P Nordmann. Chromosome-Encoded Ambler Class A -Lactamase of Kluyvera georgiana, a Probable Progenitor of a Subgroup of CTX-M Extended-Spectrum -Lactamases. *Antimicrobial agents and chemotherapy*, 46(12):4038–4040, December 2002.

[66] Rafael Cantón, José María González-Alba, and Juan Carlos Galán. CTX-M Enzymes: Origin and Diffusion. *Frontiers in microbiology*, 3:110, 2012.

[67] Søren J Sørensen, Mark Bailey, Lars H Hansen, Niels Kroer, and Stefan Wuertz. Studying plasmid horizontal transfer in situ: a critical review. *Nature Reviews Microbiology*, 3(9):700–710, September 2005.

[68] Xavier Bellanger, Hélène Guilloteau, Sébastien Bonot, and Christophe Merlin. Science of the Total Environment. *Science of the Total Environment, The*, 493(C):872–882, September 2014.

[69] I Sharon, M J Morowitz, B C Thomas, E K Costello, D A Relman, and J F Banfield. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Research*, 23(1):111–120, January 2013.

[70] D C Savage. Microbial ecology of the gastrointestinal tract. *Annual Reviews in Microbiology*, 31:107–133, 1977.

[71] Jian Xu and Jeffrey I Gordon. Honor thy symbionts. *Proceedings of the National Academy of Sciences of the United States of America*, 100(18):10452–10459, September 2003.

[72] Paul B Eckburg, Elisabeth M Bik, Charles N Bernstein, Elizabeth Purdom, Les Dethlefsen, Michael Sargent, Steven R Gill, Karen E Nelson, and David A Relman. Diversity of the human intestinal microbial flora. *Science (New York, N.Y.)*, 308(5728):1635–1638, June 2005.

[73] Junjie Qin, Ruiqiang Li, Jeroen Raes, Manimozhiyan Arumugam, Kristoffer Solvsten Burgdorf, Chaysavanh Manichanh, Trine Nielsen, Nicolas Pons, Florence Levenez, Takuji Yamada, Daniel R Mende, Junhua Li, Junming Xu, Shaochuan Li, Dongfang Li, Jianjun Cao, Bo Wang, Huiqing Liang, Huisong Zheng, Yinlong Xie, Julien Tap, Patricia Lepage, Marcelo Bertalan, Jean-Michel Batto, Torben Hansen, Denis Le Paslier, Allan Linneberg, H Bjørn Nielsen, Eric Pelletier, Pierre Renault, Thomas Sicheritz-Ponten, Keith Turner, Hongmei Zhu, Chang Yu, Shengting Li, Min Jian, Yan Zhou, Yingrui Li, Xiuqing Zhang, Songgang Li, Nan Qin, Huanming Yang, Jian Wang, Søren Brunak, Joel Dore, Francisco Guarner, Karsten Kristiansen, Oluf Pedersen, Julian Parkhill, Jean Weissenbach, MetaHIT Consortium, Peer Bork, and S Dusko Ehrlich Jun Wang. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285):59–65, April 2010.

[74] Daniel N Frank, Allison L St Amand, Robert A Feldman, Edgar C Boedeker, Noam Harpaz, and Norman R Pace. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 104(34):13780–13785, August 2007.

[75] Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402):207–214, April 2013.

[76] Peter J Turnbaugh, Micah Hamady, Tanya Yatsunenko, Brandi L Cantarel, Alexis Duncan, Ruth E Ley, Mitchell L Sogin, William J Jones, Bruce A Roe, Jason P Affourtit, Michael Egholm, Bernard Henrissat, Andrew C Heath, Rob

Knight, and Jeffrey I Gordon. A core gut microbiome in obese and lean twins. *Nature*, 457(7228):480–484, January 2009.

[77] Fredrik Bäckhed, Claire M Fraser, Yehuda Ringel, Mary Ellen Sanders, R Balfour Sartor, Philip M Sherman, James Versalovic, Vincent Young, and B Brett Finlay. Perspective. *Cell Host and Microbe*, 12(5):611–622, November 2012.

[78] Fredrik Bäckhed, Ruth E Ley, Justin L Sonnenburg, Daniel A Peterson, and Jeffrey I Gordon. Host-bacterial mutualism in the human intestine. *Science (New York, N.Y.)*, 307(5717):1915–1920, March 2005.

[79] Y K Lee and S K Mazmanian. Has the Microbiota Played a Critical Role in the Evolution of the Adaptive Immune System? *Science (New York, N.Y.)*, 330(6012):1768–1773, December 2010.

[80] Lora V Hooper, TORE MIDTVEDT, and Jeffrey I Gordon. How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annual review of nutrition*, 22:283–307, 2002.

[81] Ruth E Ley, Peter J Turnbaugh, Samuel Klein, and Jeffrey I Gordon. Microbial ecology: human gut microbes associated with obesity. *Nature*, 444(7122):1022–1023, December 2006.

[82] Peter J Turnbaugh, Ruth E Ley, Michael A Mahowald, Vincent Magrini, Elaine R Mardis, and Jeffrey I Gordon. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, 444(7122):1027–131, December 2006.

[83] Dirk Gevers, Subra Kugathasan, Lee A Denson, Yoshiki Vázquez-Baeza, Will Van Treuren, Boyu Ren, Emma Schwager, Dan Knights, Se Jin Song, Moran Yassour, Xochitl C Morgan, Aleksandar D Kostic, Chengwei Luo, Antonio González, Daniel McDonald, Yael Haberman, Thomas Walters, Susan Baker, Joel Rosh, Michael Stephens, Melvin Heyman, James Markowitz, Robert Baldassano, Anne Griffiths, Francisco Sylvester, David Mack, Sandra Kim, Wallace Crandall, Jeffrey Hyams, Curtis Huttenhower, Rob Knight, and Ramnik J Xavier. The Treatment-Naive Microbiome in New-Onset Crohn's Disease. *Cell Host and Microbe*, 15(3):382–392, March 2014.

[84] Mei Wang, Caroline Karlsson, Crister Olsson, Ingegerd Adlerberth, Agnes E Wold, David P Strachan, Paolo M Martricardi, Nils Åberg, Michael R Perkin, Salvatore Tripodi, Anthony R Coates, Bill Hesselmar, Robert Saalman, Göran Molin, and Siv Ahrné. Reduced diversity in the early fecal microbiota of infants with atopic eczema. *Journal of Allergy and Clinical Immunology*, 121(1):129–134, January 2008.

[85] Y M Sjögren, M C Jenmalm, M F Böttcher, B Björkstén, and E Sverremark-Ekström. Altered early infant gut microbiota in children developing allergy up to 5 years of age. *Clinical & Experimental Allergy*, 39(4):518–526, April 2009.

[86] Hans Bisgaard MD DMSc, Nan Li MD PhD, Klaus Bonnelykke MD PhD, Bo Lund Krogsgaard Chawes MD PhD, Thomas Skov MSc PhD, Georg Paludan-Müller MSc PhD, Jakob Stokholm MD, Birgitte Smith MD, and Karen Angeliki Krogfelt MSc PhD. Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age. *Journal of Allergy and Clinical Immunology*, 128(3):646–652.e5, September 2011.

[87] Tanya Yatsunenko, Federico E Rey, Mark J Manary, Indi Trehan, Maria Gloria Dominguez-Bello, Monica Contreras, Magda Magris, Glida Hidalgo, Robert N Baldassano, Andrey P Anokhin, Andrew C Heath, Barbara Warner, Jens Reeder, Justin Kuczynski, J Gregory Caporaso, Catherine A Lozupone, Christian Lauber, Jose Carlos Clemente, Dan Knights, Rob Knight, and Jeffrey I Gordon. Human gut microbiome viewed across age and geography. *Nature*, May 2012.

[88] J Penders, C Thijs, C Vink, F F Stelma, B Snijders, I Kummeling, P A van den Brandt, and E E Stobberingh. Factors Influencing the Composition of the Intestinal Microbiota in Early Infancy. *PEDIATRICS*, 118(2):511–521, August 2006.

[89] Maria G Dominguez-Bello, Elizabeth K Costello, Monica Contreras, Magda Magris, Glida Hidalgo, Noah Fierer, and Rob Knight. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America*, 107(26):11971–11975, 2010.

[90] Giacomo Biasucci, Belinda Benenati, Lorenzo Morelli, Elena Bessi, and Günther Boehm. Cesarean delivery may affect the early biodiversity of intestinal bacteria. *The Journal of nutrition*, 138(9):1796S–1800S, September 2008.

[91] John Penders, Cornelis Vink, Christel Driessen, Nancy London, Carel Thijs, and Ellen E Stobberingh. Quantification of Bifidobacteriumspp., Escherichia coliand Clostridium difficilein faecal samples of breast-fed and formula-fed infants by real-time PCR. *FEMS Microbiology Letters*, 243(1):141–147, February 2005.

[92] H J Harmsen, A C Wildeboer-Veloo, G C Raangs, A A Wagendorp, N Klijn, J G Bindels, and G W Welling. Analysis of intestinal flora development in breast-fed and formula-fed infants by using molecular identification and detection methods. *Journal of pediatric gastroenterology and nutrition*, 30(1):61–67, January 2000.

[93] Chana Palmer, Elisabeth M Bik, Daniel B DiGiulio, David A Relman, and Patrick O Brown. Development of the human infant intestinal microbiota. *PLoS biology*, 5(7):e177, 2007.

[94] Jeremy E Koenig, Aymé Spor, Nicholas Scalfone, Ashwana D Fricker, Jesse Stombaugh, Rob Knight, Largus T Angenent, and Ruth E Ley. Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences*, 108(Supplement 1):4578–4585, 2011.

[95] C F Favier, E E Vaughan, W M de Vos, and A D L Akkermans. Molecular Monitoring of Succession of Bacterial Communities in Human Neonates. *Applied and Environmental Microbiology*, 68(1):219–226, January 2002.

[96] Mark J Hopkins, George T Macfarlane, Elizabeth Furrie, Alemu Fite, and Sandra Macfarlane. Characterisation of intestinal bacteria in infant stools using real-time PCR and northern hybridisation analyses. *FEMS microbiology ecology*, 54(1):77–85, September 2005.

[97] H Tissier. *Recherches sur la flore intestinale des nourrissons (état normal et pathologique).* Méd.–Paris, 1900.

[98] J E L Corry, G D W Curtis, and R M Baird. *Handbook of Culture Media for Food Microbiology, Second Edition.* Progress in Industrial Microbiology. Elsevier Science, 2003.

[99] Arnold H Eggerth and Bernard H Gagnon. The bacteroides of human feces. *Journal of bacteriology*, 25(4):389, 1933.

[100] Les Dethlefsen, Paul B Eckburg, Elisabeth M Bik, and David A Relman. Assembly of the human intestinal microbiota. *Trends in Ecology & Evolution*, 21(9):517–523, September 2006.

[101] Alan W Walker, Sylvia H Duncan, Petra Louis, and Harry J Flint. Phylogeny, culturing, and metagenomics of the human gut microbiota. *Trends in microbiology*, 22(5):267–274, May 2014.

[102] M Hamady and R Knight. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Research*, 19(7):1141–1152, July 2009.

[103] M T Suzuki and S J Giovannoni. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and Environmental Microbiology*, 62(2):625–630, February 1996.

[104] M F Polz and C M Cavanaugh. Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology*, 64(10):3724–3730, October 1998.

[105] Juan M Gonzalez, Maria C Portillo, Pedro Belda-Ferre, and Alex Mira. Amplification by PCR Artificially Reduces the Proportion of the Rare Biosphere in Microbial Communities. *PloS one*, 7(1):e29973, January 2012.

[106] S Mueller, K Saunier, C Hanisch, E Norin, L Alm, T Midtvedt, A Cresci, S Silvi, C Orpianesi, M C Verdenelli, T Clavel, C Koebnick, H J F Zunft, J Doré, and M Blaut. Differences in Fecal Microbiota in Different European Study Populations in Relation to Age, Gender, and Country: a Cross-Sectional Study. *Applied and Environmental Microbiology*, 72(2):1027–1033, February 2006.

[107] A H Franks, H J Harmsen, G C Raangs, G J Jansen, F Schut, and G W Welling. Variations of bacterial populations in human feces measured by fluorescent in situ hybridization with group-specific 16S rRNA-targeted oligonucleotide probes. *Applied and Environmental Microbiology*, 64(9):3336–3345, September 1998.

[108] K H Wilson, W J Wilson, J L Radosevich, T Z DeSantis, V S Viswanathan, T A Kuczmarski, and G L Andersen. High-Density Microarray of Small-Subunit Ribosomal DNA Probes. *Applied and Environmental Microbiology*, 68(5):2535–2541, May 2002.

[109] D H Huson, A F Auch, J Qi, and S C Schuster. MEGAN analysis of metagenomic data. *Genome Research*, 17(3):377–386, February 2007.

[110] Arthur Brady and Steven L Salzberg. Phymm and PhymmBl: metagenomic phylogenetic classification with interpolated markov models. *Nature methods*, 6(9):673–676, August 2009.

[111] Kaustubh R Patil, Peter Haider, Phillip B Pope, Peter J Turnbaugh, Mark Morrison, Tobias Scheffer, and Alice C McHardy. correspondence. *Nature methods*, 8(3):191–192, March 2011.

[112] Shinichi Sunagawa, Daniel R Mende, Georg Zeller, Fernando Izquierdo-Carrasco, Simon A Berger, Jens Roat Kultima, Luis Pedro Coelho, Manimozhiyan Arumugam, Julien Tap, Henrik Bjørn Nielsen, Simon Rasmussen, Søren Brunak, Oluf Pedersen, Francisco Guarner, Willem M de Vos, Jun Wang, Junhua Li, Joel Dore, S Dusko Ehrlich, Alexandros Stamatakis, and Peer Bork. Metagenomic species profiling using universal phylogenetic marker genes. *Nature methods*, 10(12):1196–1199, October 2013.

[113] J Handelsman. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*, 68(4):669–685, December 2004.

[114] Rochellys Diaz Heijtz, Shugui Wang, Farhana Anuar, Yu Qian, Britta Björkholm, Annika Samuelsson, Martin L Hibberd, Hans Forssberg, and Sven Pettersson. Normal gut microbiota modulates brain development and behavior. *Proceedings of the National Academy of Sciences*, 108(7):3047–3052, February 2011.

[115] D J Hentges, P Pongpech, and J U Que. Hypothesis: how streptomycin treatment compromises colonisation resistance against enteric pathogens in mice. *Microbial ecology in health and disease*, 3(3):105–111, 1990.

[116] A J Fabich, S A Jones, F Z Chowdhury, A Cernosek, A Anderson, D Smalley, J W McHargue, G A Hightower, J T Smith, S M Autieri, M P Leatham, J J Lins, R L Allen, D C Laux, P S Cohen, and T Conway. Comparison of Carbon Nutrition for Pathogenic and Commensal Escherichia coli Strains in the Mouse Intestine. *Infection and immunity*, 76(3):1143–1152, February 2008.

[117] E G Zoetendal, A von Wright, T Vilpponen-Salmela, K Ben-Amor, A D L Akkermans, and W M de Vos. Mucosa-Associated Bacteria in the Human Gastrointestinal Tract Are Uniformly Distributed along the Colon and Differ from the Community Recovered from Feces. *Applied and Environmental Microbiology*, 68(7):3401–3407, July 2002.

[118] Hidenori Hayashi, Mitsuo Sakamoto, and Yoshimi Benno. Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. *Microbiology and immunology*, 46(8):535–548, 2002.

[119] A Suau, R Bonnet, M Sutren, J J Godon, G R Gibson, M D Collins, and J Doré. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Applied and Environmental Microbiology*, 65(11):4799–4807, November 1999.

[120] X Wang, S P Heazlewood, D O Krause, and T H J Florin. Molecular characterization of the microbial species that colonize human ileal and colonic mucosa by using 16S rDNA sequence analysis. *Journal of Applied Microbiology*, 95(3):508–520, September 2003.

[121] Karsten Zengler, Gerardo Toledo, Michael Rappe, James Elkins, Eric J Mathur, Jay M Short, and Martin Keller. Cultivating the uncultured. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24):15681–15686, November 2002.

[122] Karsten Zengler, Marion Walcher, Greg Clark, Imke Haller, Gerardo Toledo, Trevin Holland, Eric J Mathur, Gary Woodnutt, Jay M Short, and Martin Keller. High-Throughput Cultivation of Microorganisms Using Microcapsules. pages 124–130. Elsevier, 2005.

[123] T Kaeberlein. Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment. *Science (New York, N.Y.)*, 296(5570):1127–1129, May 2002.

[124] A Bollmann, K Lewis, and S S Epstein. Incubation of Environmental Samples in a Diffusion Chamber Increases the Diversity of Recovered Isolates. *Applied and Environmental Microbiology*, 73(20):6386–6390, October 2007.

[125] P H Janssen, P S Yates, B E Grinton, P M Taylor, and M Sait. Improved Culturability of Soil Bacteria and Isolation in Pure Culture of Novel Members of the Divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia. *Applied and Environmental Microbiology*, 68(5):2391–2396, May 2002.

[126] T Nyonyo, T Shinkai, A Tajima, and M Mitsumori. Effect of media composition, including gelling agents, on isolation of previously uncultured rumen bacteria. *Letters in Applied Microbiology*, 56(1):63–70, November 2012.

[127] Hideyuki Tamaki, Satoshi Hanada, Yuji Sekiguchi, Yasuhiro Tanaka, and Yoichi Kamagata. Effect of gelling agent on colony formation in solid cultivation of microbial community in lake sediment. *Environmental Microbiology*, 11(7):1827–1834, July 2009.

[128] J C Lagier, F Armougom, M Million, P Hugon, I Pagnier, C Robert, F Bittar, G Fournous, G Gimenez, M Maraninchi, J F Trape, E V Koonin, B La Scola, and D Raoult. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clinical Microbiology and Infection*, pages no–no, October 2012.

[129] A Belenguer, S H Duncan, A G Calder, G Holtrop, P Louis, G E Lobley, and H J Flint. Two Routes of Metabolic Cross-Feeding between Bifidobacterium adolescentis and Butyrate-Producing Anaerobes from the Human Gut. *Applied and Environmental Microbiology*, 72(5):3593–3599, May 2006.

[130] Rafael Muñoz-Tamayo, Béatrice Laroche, Éric Walter, Joel Dore, Sylvia H Duncan, Harry J Flint, and Marion Leclerc. Kinetic modelling of lactate utilization and butyrate production by key human colonic bacterial species. *FEMS microbiology ecology*, 76(3):615–624, April 2011.

[131] Shinji Fukuda, Hidehiro Toh, Koji Hase, Kenshiro Oshima, Yumiko Nakanishi, Kazutoshi Yoshimura, Toru Tobe, Julie M Clarke, David L Topping, Tohru

Suzuki, Todd D Taylor, Kikuji Itoh, Jun Kikuchi, Hidetoshi Morita, Masahira Hattori, and Hiroshi Ohno. Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature*, 469(7331):543–547, April 2012.

[132] H J SEARS, Inez Brownlee, and John K Uchiyama. Persistence of individual strains of Escherichia coli in the intestinal tract of man. *Journal of bacteriology*, 59(2):293, 1950.

[133] H J SEARS and I BROWNLEE. Further observations on the persistence of individual strains of Escherichia coli in the intestinal tract of man. *Journal of bacteriology*, 63(1):47–57, January 1952.

[134] Forough Nowrouzian, Bill Hesselmar, Robert Saalman, Inga-Lisa Strannegård, Nils Åberg, Agnes E Wold, and Ingegerd Adlerberth. Escherichia coli in Infants' Intestinal Microflora: Colonization Rate, Strain Turnover, and Virulence Gene Carriage. *Pediatric Research*, 54(1):8–14, July 2003.

[135] Forough L Nowrouzian, Agnes E Wold, and Ingegerd Adlerberth. Escherichia coli strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *The Journal of infectious diseases*, 191(7):1078–1083, April 2005.

[136] Forough L Nowrouzian, Ingegerd Adlerberth, and Agnes E Wold. Enhanced persistence in the colonic microbiota of Escherichia coli strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells. *Microbes and Infection*, 8(3):834–840, March 2006.

[137] N Karami, C Hannoun, I Adlerberth, and A E Wold. Colonization dynamics of ampicillin-resistant Escherichia coli in the infantile colonic microbiota. *Journal of Antimicrobial Chemotherapy*, 62(4):703–708, June 2008.

[138] F L Nowrouzian, O Dauwalder, H Meugnier, M Bes, J Etienne, F Vandenesch, E Lindberg, B HESSELMAR, R SAALMAN, I L Strannegard, N ABERG, I Adlerberth, A E Wold, and G Lina. Adhesin and Superantigen Genes and the Capacity of Staphylococcus aureus to Colonize the Infantile Gut. *Journal of Infectious Diseases*, 204(5):714–721, August 2011.

[139] Michael J Morowitz, Vincent J Denef, Elizabeth K Costello, Brian C Thomas, Valeriy Poroyko, David A Relman, and Jillian F Banfield. Strain-resolved community genomic analysis of gut microbial colonization in a premature infant.

Proceedings of the National Academy of Sciences, 108(3):1128–1133, January 2011.

[140] Christopher T Brown, Itai Sharon, Brian C Thomas, Cindy J Castelle, Michael J Morowitz, and Jillian F Banfield. Genome resolved analysis of a premature infant gut microbial community reveals a Varibaculum cambriense genome and a shift towards fermentation-based metabolism during the third week of life. *Microbiome*, 1(1):30, 2013.

[141] Abigail A Salyers, Anamika Gupta, and Yanping Wang. Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends in microbiology*, 12(9):412–416, September 2004.

[142] H W Smith. Transfer of antibiotic resistance from animal and human strains of Escherichia coli to resident E. coli in the alimentary tract of man. *Lancet*, 1(7607):1174–1176, June 1969.

[143] J D Anderson, W A Gillespie, and M H Richmond. Chemotherapy and antibiotic-resistance transfer between Enterobacteria in the human gastrointestinal tract. *Journal of medical microbiology*, 6(4):461–473, November 1973.

[144] W Edmund Farrar, Margene Eidson, Patricia Guerry, Stanley Falkow, Lewis M Drusin, and Richard B Roberts. Interbacterial transfer of R factor in the human intestine: in-vivo acquisition of R-factor-mediated kanamycin resistance by a multiresistant strain of Shigella sonnei. *The Journal of infectious diseases*, 126(1):27–33, 1972.

[145] V Petrocheilou, J Grinsted, and M H Richmond. R-Plasmid Transfer In Vivo in the Absence of Antibiotic Selection Pressure. *Antimicrobial agents and chemotherapy*, 10(4):753–761, October 1976.

[146] G K Morris and J G Wells. Colicin typing of Shigella sonnei. *Applied microbiology*, 27(2):312–316, February 1974.

[147] M Trobos, C H Lester, J E Olsen, N Frimodt-Moller, and A M Hammerum. Natural transfer of sulphonamide and ampicillin resistance between Escherichia coli residing in the human intestine. *Journal of Antimicrobial Chemotherapy*, 63(1):80–86, October 2008.

[148] D C Schwartz and C R Cantor. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell*, 37(1):67–75, May 1984.

[149] C H Lester, N Frimodt-Moller, T L Sorensen, D L Monnet, and A M Hammerum. In Vivo Transfer of the vanA Resistance Gene from an Enterococcus faecium Isolate of Animal Origin to an E. faecium Isolate of Human Origin in the Intestines of Human Volunteers. *Antimicrobial agents and chemotherapy*, 50(2):596–599, January 2006.

[150] Alessandra Carattoli, Alessia Bertini, Laura Villa, Vincenzo Falbo, Katie L Hopkins, and E John Threlfall. Identification of plasmids by PCR-based replicon typing. *Journal of Microbiological Methods*, 63(3):219–228, December 2005.

[151] P Bidet, B Burghoffer, V Gautier, N Brahimi, P Mariani-Kurkdjian, A El-Ghoneimi, E Bingen, and G Arlet. In Vivo Transfer of Plasmid-Encoded ACC-1 AmpC from Klebsiella pneumoniae to Escherichia coli in an Infant and Selection of Impermeability to Imipenem in K. pneumoniae. *Antimicrobial agents and chemotherapy*, 49(8):3562–3565, July 2005.

[152] N Karami, A Martner, V I Enne, S Swerkersson, I Adlerberth, and A E Wold. Transfer of an ampicillin resistance gene between two Escherichia coli strains in the bowel microbiota of an infant treated with antibiotics. *Journal of Antimicrobial Chemotherapy*, 60(5):1142–1145, September 2007.

CHAPTER 4

Transfer of multiple antibiotic resistance genes *in situ* of the infant gut without antibiotic perturbation

Heidi K. Gumpert*[1], Nahid Karami*[2], Christian Munck[1], Agnes E. Wold[2], Ingegerd Adlerberth[2], Morten O. A. Sommer[1,3]

1) Department of Systems Biology, Technical University of Denmark, DK-2800 Lyngby, Denmark.
2) Department of Infectious diseases, Section Clinical bacteriology, Sahlgrenska academy, S-405 30 Göteborg, Sweden.
3) Novo Nordisk Foundation Center for Biosustainability, DK-2900 Hørsholm, Denmark.

4.1 Abstract

The human gut is one of the densest microbial ecosystems and is believed to play an important role in the exchange of antibiotic resistance genes. We study the dynamics of co-existing *Escherichia coli* lineages in a longitudinally followed infant not receiving antibiotics. Using whole genome sequencing, we capture the exchange of multidrug resistance genes and identify a clinically relevant conjugative plasmid mediating the transfer. Quantification of the co-existing lineages reveals that the resistant transconjugant is maintained for months, demonstrating that antibiotic resistance genes disseminate and remain in the gut microbiome even in absence of antibiotic selection. Furthermore, we observe large genomic deletions, as well as acquisition of phages in these lineages during their colonization of the human gut. Our findings highlight the dynamic nature of the human gut microbiota and provide the first genomic description of antibiotic resistance gene transfer between bacteria in the unperturbed human gut.

4.2 Introduction

Resistance to antibiotics is an ever-growing public health concern. Antibiotic resistance in bacteria can either be achieved through mutation in the genome or via horizontal acquisition of foreign genetic material that confers resistance [1]. While antibiotic resistance achieved through mutation is of significant importance during long term treatment of chronic infections [2], horizontal gene transfer plays a key role in the ever-growing problems with multidrug resistant human pathogens [3].

The environment is considered a reservoir of antibiotic resistance genes [4], and indirect evidence has shown that antibiotic resistance genes have disseminated between the soil microbiota to both commensal intestinal microbes and to clinical pathogens [5, 6]. In addition to the flow of genetic material between environmental and human resevoirs, antibiotic resistance genes have also been suggested to spread between bacteria within the human microbiome [7]. In the human intestine, this transfer would not only be between members of the commensal microbiota, but would also include non-commensal bacteria transiting through the intestine [8]. While the gut microbiome has been the subject of numerous metagenomic studies [9, 10, 11], including a recent study constructing complete genomes of various species and strains from metagenomic data [12], the use of metagenomics is not well-suited to detect

horizontal gene transfer events between species or strains in the human gut.

Transfer of antibiotic resistance genes within the gut microbiota can occur in animals [13], as well as in humans [14, 15]. These studies have directly demonstrated antibiotic resistance transfer between bacteria in the gut. However, these studies set up an artificial scenario where the host is fed a strain that can donate antibiotic resistance genes to the commensal microbiota. A recipient strain or the commensal bacteria is subsequently monitored to detect if the antibiotic resistance genes have been transferred from the donor strain. Despite employing an artificial experiment set-up, studies have started to elucidate factors known to influence transfer between bacteria. For example, Stecher *et al.* reported that pathogen-induced gut inflammation of lab mice gives rise to an environment where lineages of various *Enterobacteriaceae* species can bloom resulting in unprecedented rates of horizontal gene transfer between these bacteria [16].

A few instances of natural transfer of antibiotic resistance genes in the unperturbed human gut microbiota have been published and these reports describe changes in the antibiotic resistance profiles of strains collected from infants undergoing antibiotic treatment [17, 18]. Additionally, a retrospective study examining a collection of *Bacteroides* collected over a period of 40 years demonstrated that extensive resistance gene exchange has occurred between species of *Bacteroides* and other genera in the human colon [19]. Even though the gut is considered a reservoir of antibiotics resistance genes and considered to serve as hub for horizontal gene exchange, full genomic data for strains exchanging antibiotic resistance genes in situ of the human gut has yet to be reported. In this study, we characterize the transfer of antibiotic resistance between co-existing *Escherichia coli* lineages in the infant gut at the genomic level using whole genome sequencing. As this transfer occurred in the absence of antibiotic treatment and the resistant transconjugant was maintained for months in the gut, we believe this study provides evidence that resistance genes are stably maintained once acquired.

4.3 Results

Study Material

Our study material was selected from an infant enrolled in the ALLERGYFLORA study, which was designed to examine the link between the infant gut microbial colonization pattern and the development of allergies [20]. Faecal samples from this

infant were obtained over the course of the first year of life at 3 days, 1, 2 and 4 weeks, and 2, 6 and 12 months. The samples were quantitatively cultured for *Enterobacteriaceae* selection. Isolates identified as *E. coli* by biotyping were initially assigned to specific lineages based on random amplified polymorphic DNA (RAPD) typing. The sensitivity threshold of the culture method was $10^{2.5}$ colony-forming units (CFU) per gram of fecal matter [21], which is several orders of magnitude more sensitive than current metagenomic sequencing studies [12].

The antibiotic resistance profile was determined for each isolate. Interestingly, a change in the antibiotic resistance profile was detected in a set of consecutive isolates from one lineage colonizing one of the infants enrolled in the ALLERGYFLORA study. The set of *E. coli* lineages collected from this infant are the subject of this study, chosen to investigate the genome dynamics involved in the acquisition and persistence of antibiotic resistance.

E. coli Lineage Sampling

E. coli isolates were obtained from the infant at sampling time points from two weeks to twelve months after birth (Fig. 4.1). In total, three distinct lineages were identified: A, B and C. The sampling at two and four weeks after birth yielded only isolates belonging to lineage A, which were sensitive to all antibiotics tested (Fig. 4.1, Supplementary Table 4.1). At the 2-month sampling time, lineage B was observed. Lineage B was resistant to the antibiotics ampicillin, piperacillin, streptomycin and sulfamethoxazole (Fig. 4.1). At the two-month sampling time, the antibiotic resistance profile of lineage A changed. Lineage A was now resistant to ampicillin, piperacillin, streptomycin and sulfamethoxazole. Notably, the change in the resistance profile of lineage A at two months matches that of the incoming lineage B suggesting a transfer of resistance determinants from lineage B to A. Lineages A and B were both present at the six months sampling time with no changes in the antibiotic resistance profile. However, at the 12-month sampling time only lineage B was still present, with the addition of lineage C, which is only resistant to sulfamethoxazole (Fig. 4.1).

Genome Sequencing of the Isolates

A total of 13 isolates from lineages A, B and C were genome sequenced. At least one isolate sequenced per lineage per sampling point when present was selected (Supple-

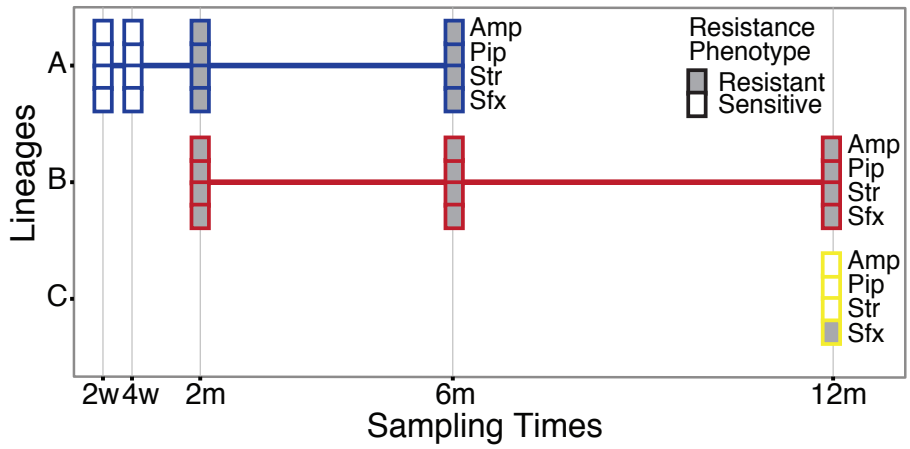


Figure 4.1 | Sampling and antibiotic resistance of the *E. coli* strains. A total of three *E. coli* lineages were sampled from the infant's intestinal microbiota over the course of the first year of life. Stacked squares indicate both the presence of the lineage at the sampling points, in addition to their antibiotic resistance profile to ampicillin (Amp), piperacillin (Pip), streptomycin (Str), and sulfamethoxazole (Sfx) at the time points. Filled coloured squares indicate resistant isolates, and non-filled squares indicate sensitive isolates (Supplementary Table 4.1 for MICs).

mentary Table 4.2). Lineage A isolates included two from the 2 week sampling time, $2w_1$ and $2w_2$, one from 4 weeks and 2 months, 4w and 2m, respectively, and two from 6 months, $6m_1$ and $6m_2$. Lineage B isolates included two isolates form 2 months, $2m_1$ and $2m_2$, one from 6 months, 6m, and two from 12 months, $12m_1$ and $12m_2$. Lineage C isolates included two from 12 months, $12m_1$ and $12m_2$. Reads from each isolate were assembled [22] and annotated via RAST [23]. Comparisons between isolates were made based on mapping reads from one isolate to assembled contigs from another isolate [24] and identifying single nucleotide polymorphisms (SNPs) [25] and assessing read coverage [26], as well as by aligning contigs [27].

To confirm lineage identities of the isolates, we assessed both the number of SNPs and the amount of total genomic content shared between lineages. One of the first sampled isolates from each lineage was selected for comparison, specifically lineage A $2w_2$, lineage B $2m_2$, and lineage C $12m_2$. Both lineages A and C have on the order of 90,000 SNPs when compared to lineage C (Supplementary Table 4.3). Interestingly,

lineages A and C have an order of magnitude fewer SNPs when compared to each other, on the order of 7,000 SNPs, than when compared to B. Similarly when comparing the percentage of the genomic content shared between the lineages, lineages A and C shared between 79.7% and 82.3% in common with lineage B (Supplementary Table 4.4). Lineage A shared 93.6% of the genomic content from lineage C, and lineage C shared 95.5% from lineage A. While these results indicate that lineages A and C are more closely related to each other than to the lineage B, the amount of SNPs and differences in genomic content reveal that they are different lineages. RAPD-typing of the isolates was sensitive enough to successfully classify the colonies into the three distinct lineages.

Evolutionary relationships amongst the isolates within a lineage were established based on the number of SNPs identified by aligning reads to one of the first sequenced isolates, where again lineage A $2w_2$, lineage B $2m_2$ and lineage C $12m_2$ was selected (Supplementary Tables 4.5, 4.6, 4.7 for full list of SNPs with annotations for lineages A, B and C, respectively). SNPs identified in isolates from lineages A, B and C produced consistent phylogenetic trees and show a progression in the acquisition of SNPs (Fig. 4.2). In the A lineage, one SNP was identified between the $2m_1$+$2m_2$ and the 4w isolates, and this SNP was again identified in all isolates of lineage A subsequently sampled. Isolates 2m, $6m_1$ and $6m_2$ all branched out after this point. Similarly, lineage B had one SNP propagated to all isolates subsequently obtained after the initial SNP detection. The 6m isolate branched out after this point, while the 12m isolates share 4 SNPs in common before branching. One SNP was identified between the two lineage C 12m isolates. The consistent phylogenetic tree and the progression of SNPs indicate that the isolates sampled are representative clones for both the A and B lineages.

To get further information about the isolated lineages we compared the isolates to previously sequenced *E. coli* genomes found in the NCBI reference sequence database, based on genomic content in common and number of SNPs. Lineage A had 93.2% genomic content in common with the asymptomatic bacteriuria (ABU) strain, ABU 83972 (NC_017631.1) [28]. ABU 83972 was also the closest previously sequenced strain to lineage C as well, sharing 91.5% genomic content. The ABU 83972 strain, which is closely related to the pathogenic CFT073 *E. coli* strain, has been reported to show a tendency to evolve toward commensalism from virulence during asymptomatic colonization of the bladder via the loss of gene function [28]. Lineage B shared 95.4% genomic content with UMN026 (NC_011751.1), an extra-intestinal pathogenic strain

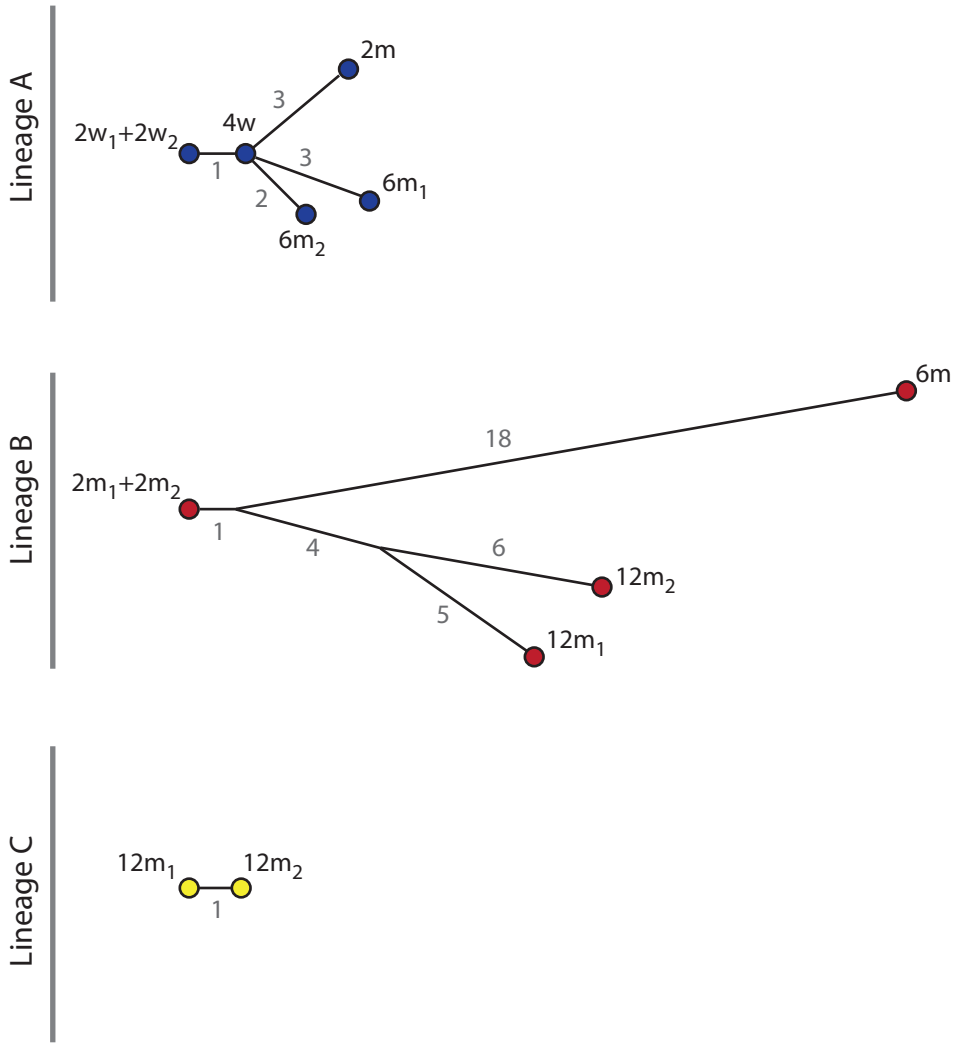


Figure 4.2 | Phylogenetic tree of the isolates. Phylogenetic trees based on the number of SNPs found in each of the isolates of lineages A, B and C. The grey value next to each branch indicates the number of SNPs between isolates.

[29]. The UMN026 strain belongs to clonal group A, which is a globally spread uropathogenic clone that frequently carry multiple antibiotic resistance determinants [30]. Lineage B was also determined to belong to the clonal group A via gene specific PCR [31].

Additionally, we also determined the multi locus sequence types (MLST) and the phylogenetic groups of the lineages. Surprisingly, lineages A and C belonged to the same MLST group, ST12, (Supplementary Table 4.8), yet RAPD typing was able to distinguish between these lineages. Lineage B belonged to ST782. Lineages A and C belonged to phylogenetic group B2, and lineage B belonged to group D.

Lineage genome dynamics

By examining the genomic changes in the isolates, we observed that the lineages were highly dynamic. Genome dynamics included acquisition of antibiotic resistance, major genomic deletions, plasmid loss, and phage integrations (Fig. 4.3).

To identify the genomic changes underlying the acquisition of antibiotic resistance in lineage A, sequence data collected from the sensitive isolates ($2w_1$, $2w_2$ and 4w) were compared to sequence data from the resistant isolates (2m, $6m_1$ and $6m_2$). Two non-conservative genomic mutations in the betaine aldehyde dehydrogenase (*betB*) and phosphoenolpyruvate carboxylase (*pckA*) genes were identified, however, these would not be expected to contribute to antibiotic resistance. Instead, additional genetic information totaling 90kb was found in the resistant isolates from lineage A that were collected at the two and six month sampling points compared to lineage A isolates from the 2 and 4 week time points (Fig. 4.1). The read coverage of the newly acquired genetic information had 2 times greater coverage depth compared to that of the genome, suggesting that a newly acquired low-copy number plasmid harbored the antibiotic resistance determinants. This putative plasmid contained conjugative transfer genes (*trb*, *tra* and *pil* operons). Additionally, the putative plasmid contained a beta-lactamase (bla_{TEM-1c}), which has been reported to confer resistance to penicillins, including ampicillin [32]. The TEM-1 beta-lactamases are widely disseminated amongst several different pathogens including *E. coli*, *Haemophilus influenza* and *Neisseria gonorrhoeae* [32]. Furthermore, two genes, aminoglycoside 3'-phosphotransferase (*strA*) and streptomycin phosphotransferase (*strB*), known to mediate resistance to streptomycin and other aminoglycosides, were identified, the latter being widely disseminated on a broad host range plasmid [33]. Lastly, a dihy-

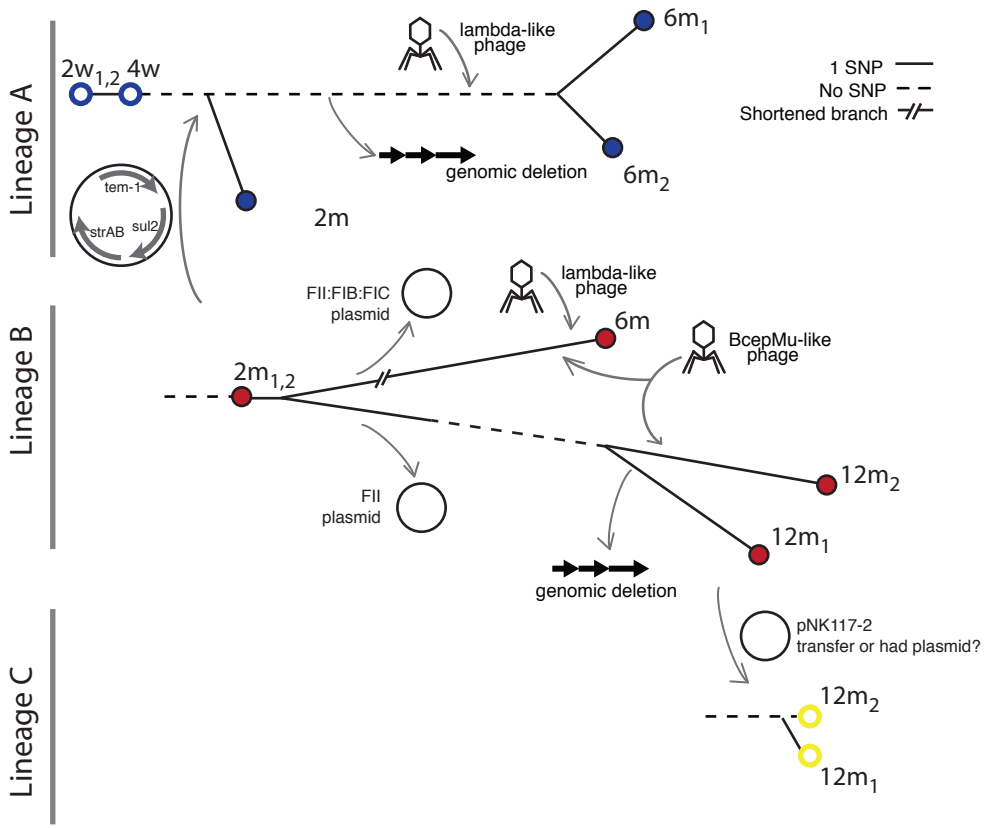


Figure 4.3 | Overview of lineage genome dynamics. The transfer of a multidrug resistance plasmid from lineage B to lineage A occurred before the 2 month sampling time. The transfer occurred before diversification of lineage A. A large genomic deletion occurred in lineage A after 2 months but before the 6 month sampling point, as detected in the isolates $6m_1$ and $6m_2$ isolated at 6 months. At the 6 month sampling point, both lineages A and B were infected by different lambda-like phages. A BcepMu-like phage infecting lineage B was detected in 6m and $12m_2$ isolates, identified in different regions of the chromosome indicating that two distinct phage integrations occurred. Lineage B contained two plasmids containing a FII replicon in the 2m isolates, whereas isolate 6m contained only the plasmid containing the FII replicon, and the 12m isolates contained the plasmid with the FII:FIB:FIC replicons. A genomic deletion was detected in lineage B $12m_1$ isolate.

Figure 4.3 continued | No isolates of lineage A were obtained at the final sampling time at 12 months, but a new isolate from lineage C is sampled along with lineage B. Both lineage C $12m_1$ and $12m_2$ isolates contained a plasmid, pNK117-2, that was 100% identical to a plasmid found in lineage B. This could be due to another conjugative transfer event, or lineage C could simply have harboured the pNK117-2 on entry in the gut.

dropteroate synthase gene (*sul2*) was found on the putative plasmid. The *sul2* gene is known to mediate resistance to sulfonamides and is also frequently found in clinical isolates [34].

Querying sequence databases for similar plasmids yielded the clinically important conjugative, IncI1-type pHUSEC41-1 plasmid of 91,942bp [35]. Aligning contigs from the isolates in this study to pHUSEC41-1 resulted in 99.3% coverage of the plasmid with an average of 99.0% identity. The alignment also showed that there were no insertions in the transferred plasmid compared to pHUSEC41-1. The pHUSEC41-1 plasmid was initially identified in the *E. coli* serotype O104:H4 strain HUSEC41 isolated from a child in Germany with hemolytic-uremic syndrome (HUS) [35]. The plasmid has additionally been found in other sequenced *E. coli* isolates of serotype O104:H4 isolated from patients in France with sporadic cases of HUS [36]. Here, the plasmid is found in *E. coli* strains of serotypes O73: K-:H18 and O179: K12:H-isolated from a Swedish infant, highlighting the wide dissemination of this multiple antibiotic resistance plasmid amongst various *E. coli* strains.

The phenotypic resistance patterns (Fig. 4.1) suggested that the horizontally acquired resistance was transferred from lineage B to lineage A. To assess this we analyzed lineage B to see if it contained the acquired genetic information of strain A. Aligning reads from lineage B to the acquired genetic information of lineage A resulted in coverage of 100% of the new genetic information with only one identified SNP variant (Fig. 4.4). The high-degree of identity between the plasmids and the change in the antibiotic resistance profile of lineage A to match the profile of lineage B is consistent with strain B transferring its antibiotic resistance plasmid to lineage A.

Plasmid content was also found to vary in lineage B, but here plasmid loss was discovered. We used the PlasmidFinder tool to identify possible plasmid replicons (Supplementary Table 4.9) [37]. In the $2m_1$ and $2m_2$ isolates, multiple IncF plasmid

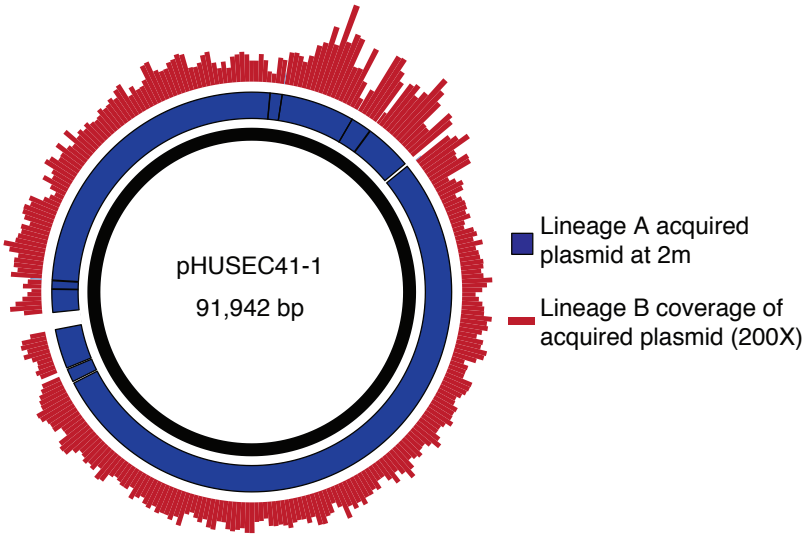


Figure 4.4 | Transfer of a plasmid mediating antibiotic resistance. Contigs corresponding to the newly acquired plasmid were identified by analyzing differences in the read alignment coverage before and after the change in the resistance profile. Reads from lineage B are mapped to the acquired plasmid contigs of A, displaying coverage depth. High coverage and identity between the strains was observed. The acquired plasmid contigs of A were aligned to the sequence of pHUSEC41-1.

replicons were identified, namely two FII, one FIB, and one FIC. The two FII replicons were distinct and identified on different contigs. In the 6m isolate, only one of the FII replicons was identified, while the other FII replicon and the FIB and FIC replicons were found in the $12m_1$ and $12m_2$ isolates. The FII and FIC replicons were identified on the same contig for both $12m_1$ and $12m_2$ assembled genomes. While we were unable to close these plasmids, evidence that the FIB replicon is also present on the FII:FIC plasmid is the pattern of its presence, ie. the FIB is observed in both $12m_1$ and $12m_2$ and not in 6m isolates. Typing of these plasmids using the IncF pMLST typing scheme yielded only a known FIB allele, B:6 [38, 37]. Here, we observe a reduction in the plasmid content of lineage B of plasmids with similar replicons.

Both FII and FII:FIB:FIB plasmids shared high homology over the 32kb regin encoding for IncF conjugation machinery. Contigs identified to be a part of the FII

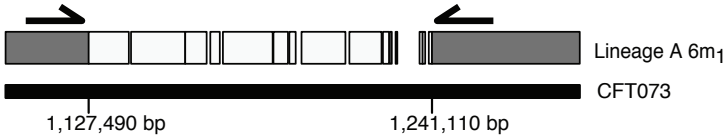


Figure 4.5 | Large deletion in the genome of lineage A. Contigs from strain A were aligned to reference genome CFT073. Dark grey colored contigs represent regions flanking the excision. Uncoloured contigs represent the region lost due to the deletion. Arrows indicate the position of the primers designed based on the CFT073 genome used to confirm the genomic excision.

plasmid from the 2m and 6m isolates totaled 54 kb, and contained genes encoding for microcin B17 (Supplementary Table 4.10). Contigs identified to be part of the FII:FIB:FIC plasmid from the 2m and 12m isolates totaled 110 kb. Genes involved in iron transport were identified via RAST (Supplementary Table 4.11) [23]. A 20 kb region on the contig with the FIB replicon was dedicated to genes encoding for heavy metal resistance.

Interestingly, a large deletion was detected in lineage A isolates at the six-month sampling point, i.e. after acquisition of the resistance plasmid. The lost contigs, totaling 83kb, aligned to a contiguous region in a similar, sequenced strain, *E. coli* CFT 073 (Fig. 4.5) (NC_004431). The ABU 83972 strain, which has high homology to the CFT 073 strain, was not used due to a documented prophage integrated in this region [28]. Using the alignment information, PCR assays were conducted to establish that the deletion was an excision (Fig. 4.5, Supplementary Fig. 4.1).

Annotated genes located in the deleted segment included iron scavenging genes, such as the *iroA* gene cluster and the hemolysin activator protein, peptide antibiotic genes microcin H47 and colicin-E1, which target gram-negative bacteria in general and *E. coli* specifically, respectively, and antigen 43, which may have a role in adhesion [39]. Lastly, genes involved in fatty-acid synthesis, carbohydrate and amino acid metabolism were also lost as a result of the deletion (Supplementary Table 4.12 for complete list).

Similarly, a large deletion was also identified in lineage B, but only in the $12m_1$

isolate (Fig 4.3). The contigs associated with this deletion aligned to a contiguous region in the UMN026 chromosome, spanning a region of 26 kb. This region included *pap* genes encoded for P fimbriae (class II), mobile elements and a TonB-dependent receptor (Supplementary Table 4.13 for complete list). While P fimbriae are a well-known adhesin involved in the pathogenesis of urinary tract infections [40], they have also been linked to enhanced persistence of resident *E. coli* strains in the gut microbiota in infants [41, 20, 42].

Furthermore, phage content varied in the lineages across sampling time points, in addition to the several stably integrated phages. At the 6-month sampling time, a phage bloom had occurred where both lineage A and B were infected by phages (Fig. 4.3). Lineage A was infected by a 27.1 kilobase (kb) Lambda-like phage, while lineage B was infected by two phages: a different Lambda-like phage of 21.6kb, and a BcepMu-like phage of 45.4kb. Notably, the BcepMu-like phage infecting lineage B had very high similarity to an integrated phage in the genome of lineage A (at least 99.9% similarity covering 97.7%), meaning that lineage B had just acquired an extra 40kb of genetic material on par with lineage A. However, the phage likely did not originate from lineage A due to the differences in both sequence and structure in one region. The integration of several phages at this time point comprising more than 90kb highlights how phages play a vital role in horizontal gene transfer. Interestingly, only the BcepMu-like phage from lineage B was identified in an isolate again from a later sampling time, and only in the $12m_2$ isolate. Aligning the flanking regions of the phage to the $2m_2$ isolate revealed that the phage integrated into two different regions in the genome. This suggests that two separate phage integration events occurred in lineage B. With regards to phylogenetic tree where $12m_1$ and $12m_2$ isolates share 4 SNPs in common after branching from the 6m isolate (Fig. 4.2), this also would appear to be the most likely explanation how both 6m and $12m_2$ isolates contain the BcepMu-like prophage.

At the last sampling time at 12 months, lineage C was sampled for the first time. Comparing the genome of lineage C to lineages A and B revealed that it was most related to lineage A. However, lineage C contained one plasmid and this plasmid had 100% similarity to a plasmid from lineage B (Fig. 4.3). This 35.8 kb plasmid, termed pNK117-2, contains the *virB* conjugation system (Supplementary Table 4.14). A blastn search revealed that pRPEC180_47 was the most closely related plasmid with 85% coverage and 98% similarity [43]. The pRPEC180_47 plasmid belongs to the IncX1 incompatibility group. The PlasmidFinder web service did not identify

any replicons in lineage C [37] (Supplementary Table 4.9). The grouping for IncX1 plasmids is based on the *bis* gene encoding an auxiliary replication initiation protein from the pOLA52 plasmid [37], which the pNK117-2 plasmid was found to only the first 94 bp (20%) of *bis*. However, the plasmid does bear resemblance to other IncX1 plasmids (Fig 4.6). Typically, IncX plasmids follow the generalized structure of *pir-bis-par-hns-topB-pilX-actX-taxCA* [43]. A homologous *bis* gene was found on an NDM-1 producing plasmid from *Raoultella planticola* RJA274. Additionally, homologues for par, *hns*, *topB*,*pilX* (*virB*), *actX*, and *taxCA* from pOLA52 were identified on pNK117-2 and in the same order (Fig 4.6).

The presence of pNK117-2 in both lineage B and C with 100% sequence identity could indicate that another conjugative transfer event occurred *in situ* of the gut. However, lineage C was only sampled at 12 months, so we did not sample lineage C at a previous sampling time without pNK117-2. While the presence of both lineage B and C harbouring pNK117-2 in the gut may seem like a transfer event is the most likely reason they both contain pNK117-2, there are examples of sequenced plasmids with very high identity to other previously sequenced plasmids in different genetic backgrounds (eg. the highly-similar pHUSEC41-1 in this paper, and the pUTI89-like plasmids [44, 45, 46, 47]).

E. coli lineage population counts in relation to genome dynamics

To investigate whether the various genomic events occurring in the *E. coli* lineages possibly affected their fitness in the gut microbiota, we examined the population counts of lineages A and B at the different sampling points (Fig. 4.7). In general, population numbers of *E. coli* decrease in the gut of infants over the first year of life, in parallel with the establishment of a microbiota dominated by anaerobic bacteria [20].

Interestingly, the acquisition of the resistance plasmid in lineage A was associated with a very steep drop in population counts, from $10^{10.2}$CFU per gram of fecal matter in the 4 week sample to $10^{7.8}$per gram in the 2 month sample. This could possibly relate to a fitness cost imposed on lineage A from carrying the resistance plasmid. This possibility was further supported by pair-wise in vitro growth competition experiments comparing the growth of a lineage A isolate from the 4 week sampling, before acquisition of the plasmid, and a lineage A isolate from the 2 month sampling, after acquisition of the plasmid. In these experiments carriage of the plasmid incurred

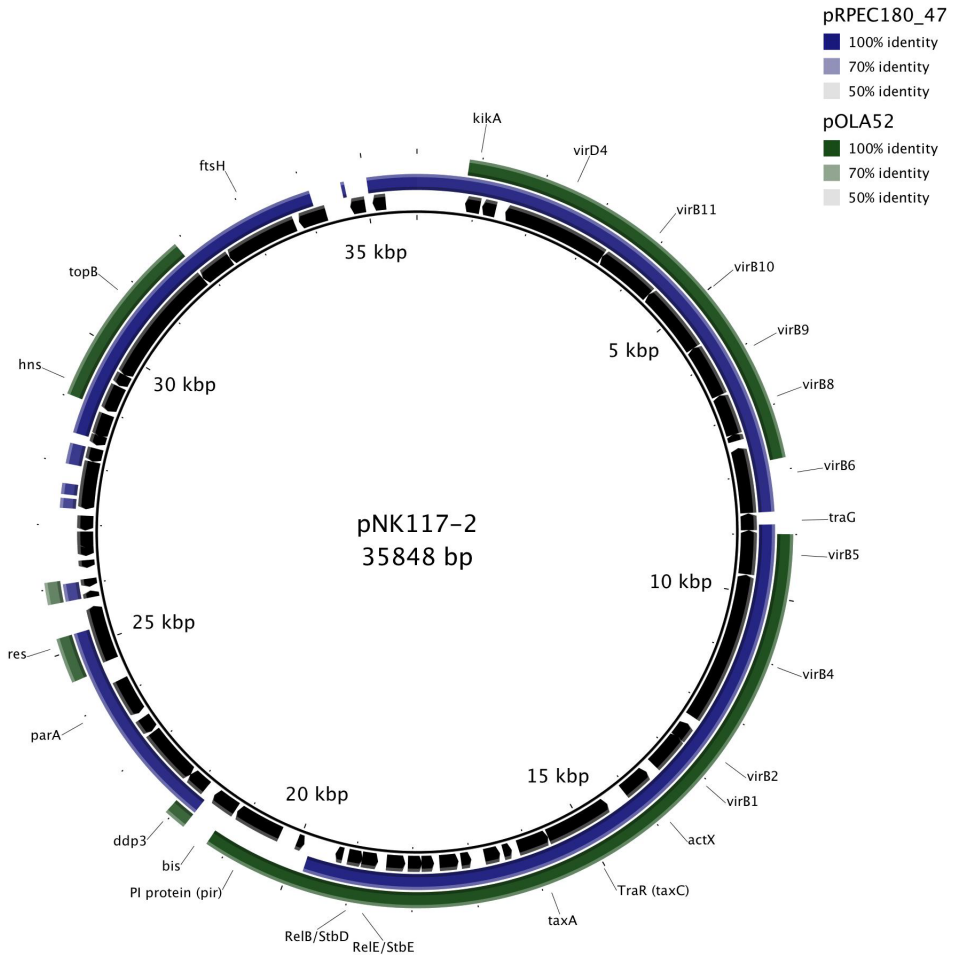


Figure 4.6 | Plasmid NK117-2 identified in both lineage B and C. Here it is compared to IncX1 plasmids pRPEC180_47 (middle ring, blue) and pOLA52 (outer ring, green). Open reading frames (ORFs) identified on pNK117-2 are drawn in the inner most ring in black, with arrows indicating the reading direction. Annotations for some ORFs are labeled outside of the rings.

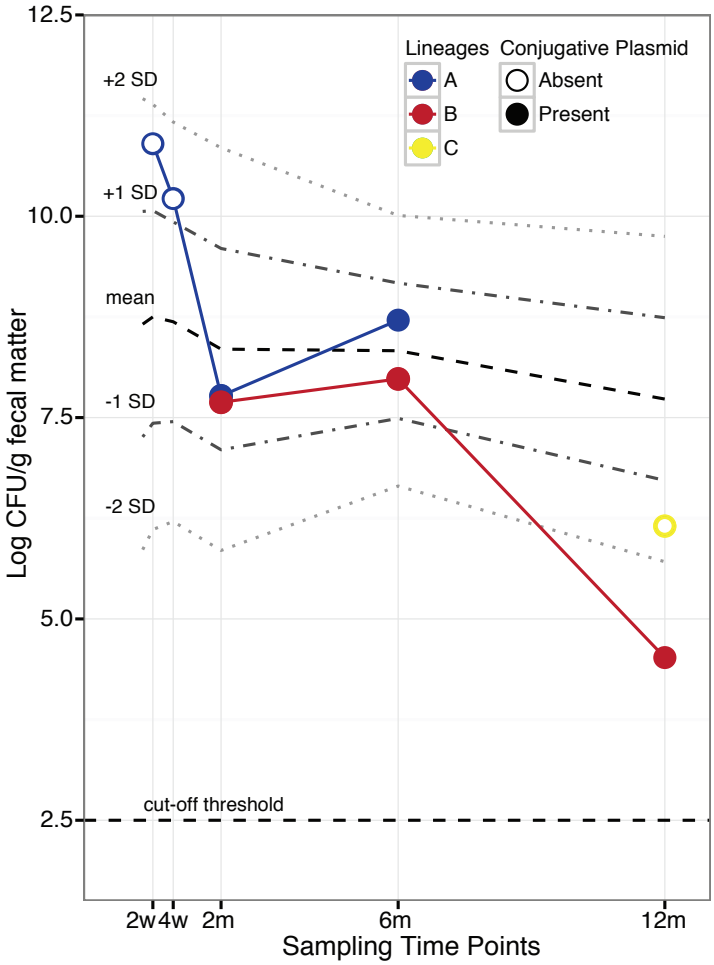


Figure 4.7 | Population counts of co-existing E. coli lineages. Faecal population counts of E. coli lineages A, B and C at different sampling points during the first year of life of the infant studied. For comparison, the mean population levels and +/- 1 and 2 standard deviations (SD) at the same sampling points for 272 E. coli strains isolated from 128 infants in the ALLERGYFLORA cohort are indicated in the figure.

a cost of -6.3% +/- 1.9% per generation on lineage A. However, despite the fitness costs the lineage persisted in the gut for at least another 4 months, and even increased in numbers during this time.

At the 2 month sampling time, when lineage B was first sampled, lineages A and B had roughly the same population counts, $10^{7.8}$ and $10^{7.7}$ CFU/g, respectively. At the 6 month sampling time, the population counts of lineage A were several times higher than the counts of lineage B ($10^{8.7}$ versus $10^{8.0}$ CFU/g,) and had also increased almost one log since the 2 month sampling point. The large genomic deletion in the A lineage was first detected at the 6 month sampling point, as well as the phage infections in both the A and B lineages. Possibly, the large deletion occurring in lineage A between the 2 and 6 month sampling points reduced the fitness cost imposed on lineage A by the resistance plasmid. It is also possible that the phage infections differentially affected lineage A and B, providing lineage A with a competitive advantage.

4.4 Discussion

The human gut, as a hub for horizontal gene exchange, is expected to play an important role in the exchange of antibiotic resistance genes. Yet to our knowledge, there are no prior reports documenting horizontal gene exchange events at the genomic level between bacteria in the unperturbed human gut. Here, by genome sequencing a collection of co-existing *E. coli* strains, we describe several cases of horizontal gene transfer mediated by both integrative phages and conjugative plasmids. The substantial genome dynamics captured in this study highlight the highly dynamic nature of the gut microbiota.

Specifically, we document at the genomic level transfer of a multiple drug resistance plasmid between co-existing bacterial lineages in the human gut. Notably, the transfer occurred in the gut of an infant not treated with antibiotics. This provides compelling evidence to support the idea of the human gut serving as a hotspot for horizontal resistance gene exchange even in absence of selective pressure from antibiotics. Furthermore, the A lineage persisted in the infant gut for months after acquisition of the resistance plasmid. The fact that the transconjugant survived and even increased its population counts highlights that resistance genes might easily disseminate also in healthy individuals never treated with antibiotics.

We also identify the entire mobile genetic element responsible for this horizontal gene transfer and discover that it is closely related to the clinically relevant multidrug

resistance plasmid pHUSEC41-1. This study highlights the advantages of studying the longitudinal dynamics of co-existing bacterial lineages in the gut microbiota as a complement to metagenomic sequencing studies. The power of this approach is expected to increase as cultivation methods for representatively sampling the gut microbiota further improves and we anticipate that further studies augmenting metagenomic sequencing with genomic sequencing will provide a richer and more detailed view of the highly dynamic nature of strains in the human gut microbiota.

4.5 Materials and Methods

Strain isolation and population counts

Fecal samples were obtained from an infant enrolled in the ALLERGYFLORA study [20]. A sample of the rectal flora was obtained using a cotton-tipped swab at 3 days after birth. The infant?s parents collected fecal samples at 1, 2, and 4 weeks, and 2, 6 and 12 months of age. Samples were placed in airtight containers with an anaerobic sachet (AnaeroGen Compact, Oxoid Ltd, UK) and refrigerated until brought to the lab, where they were serially diluted and cultivated within 24 hours on a range of non-selective and selective media, including Drigalski agar plates [48] for the isolation of *Enterobacteriaceae*. The detection limit was $10^{2.5}$ CFU/g fecal matter. Isolates resembling Enterobacteriaceae were selected from each positive culture based on differing colony morphological characteristics such as size, shape, color or mucoid appearance and speciated using API 20E biotyping (API Systems SA, France). Each morphotype was enumerated separately, and strain identities of the enumerated morphotypes were confirmed using random amplified polymorphic DNA (RAPD) typing [20]. Initial confirmation of the RAPD-typing was confirmed by performing pulsed-field gel electrophoresis (PFGE). Isolated strains were subjected to complete serotyping (O:K:H) (Statens Serum Institute, Denmark). The sample taken at 3 days and 1 week after birth did not yield any *E. coli* strains. In samples taken from 2 weeks to one year after birth, a total of 3 *E. coli* lineages were isolated from colonies. From these 5 sampling times positive for *E. coli*, a total of 13 isolates were selected and stored for further analysis.

Antibiotic susceptibility and Minimum Inhibitory Concentration (MIC) determination

While isolating *E. coli* strains from fecal samples, isolates were also tested for their susceptibility to the following antibiotics using the disc diffusion method (Oxoid, Sweden): ampicillin, amoxicillin/clavulanic acid, piperacillin, mecillinam, cefadroxil, ceftazidime, cefuroxime, cefoxitin, chloramphenicol, gentamicin, tobramycin, streptomycin, nitrofurantoin, nalidixic acid, tetracycline, trimethoprim and sulphonamide. From the saved isolates, the MIC of the following isolates were determined, such that there was one isolate selected per lineage per sampling point: lineage A $2w_1$, 4w, 2m, $6m_1$ isolates, lineage B $2m_1$, 6m, $12m_1$ isolates, and lineage C $12m_1$ isolate. The MICs of these 8 isolates were performed in 96 well micro-titer plates using the following antibiotics: ampicillin, piperacillin, mecillinam, ceftazidime, cefuroxime, cefoxitin, chloramphenicol, gentamycin, tobramycin, streptomycin, nalidixic acid, tetracycline, trimethoprim, sulfamethoxazole (Sigma). Each drug gradient consisted of 11 points in a two-fold dilution series prepared in Mueller Hinton broth II, cation-adjusted (MHBII) (Sigma) with a total of 150 μl in each well. The MIC plates were inoculated from diluted over-night cultures grown in MHBII, such that approximately 10^5 cells were added per well using a 96-pin replicator. The plates were incubated at 37°Cfor 18 - 20 hours and the optical density (OD) at 600 nm was read on a BioTek Epoch plate reader.

PCR for determination of phylogenetic group identity and clonal group

Strains were also characterized with respect to phylogenetic group identity (A, B1, B2 or D) by triplex PCR using primers detecting the genes *chuA* and *yjaA* and an anonymous DNA fragment designated TSPE4.C2 [49]. A PCR recognizing members of the clonal group A (CGA), which is characteristic of a globally spread, pyelonephritogenic clone, was applied using previously described conditions [31].

Genome Sequencing

Genomic DNA from each of the 13 isolates was obtained using an UltraClean® Microbial DNA Isolation Kit (Mobio Laboratories, Inc.). Both single-end (SE) read libraries and paired-end (PE) read libraries were prepared (Supplementary Table 4.2).

For SE library preparation, the extracted DNA was sheared into 200bp fragments using a Covaris E210 and barcoded libraries were constructed for Illumina sequencing and performed by Partners HealthCare Center for Personalized Genetic Medicine (Massachusetts, USA). For PE library preparation, the TruSeq Nano DNA kit was used (Illumina, USA) and MiSeq sequencing was performed at the Novo Nordisk Foundation Centre for Biosustainability (Denmark).

Sequence Analysis

Reads from each isolate was assembled using Velvet [22]. Single-end and paired-end reads were assembled with variable k-mer sizes (Supplementary Table 4.2). Contigs with less than 500 bp were filtered. Contigs were corrected by aligning reads using Bowtie2 (version 2.1.0) [24], calling single-nucleotide polymorphisms (SNPs) using SAMtools (version 0.1.19) [25], and edited using custom biopython scripts [50]. Contigs were annotated using the RAST server [23]. Bowtie2 and SAMtools were also used to determine the number of SNPs between isolates. Identified variants that had a Phred quality score less than 50 or with less than 90% of the high-quality reads as the variant were filtered. SNPs occurring in short homologous regions after genomic deletions or acquisitions were also filtered. BEDtools [26] was used to calculate read coverage across genomes and thus identify acquired or deleted genomic information. MUMmer was used to align sequences [51].

Multi locus sequence typing (MLST)

MLST groups were determined based on the following genes extracted from the assembled contigs: *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*, using the database hosted at http://mlst.warwick.ac.uk/mlst/dbs/Ecoli [52].

Phage Identification

The PHAST phage search tool server [53] was used to identify possible intact phages in the contigs. In addition, BLAST was used to identify similar previously described phages. Phage integration sites were determined by aligning contigs containing the flanking regions of the phage to an earlier isolate not containing the prophage.

Plasmid Analysis

The PlasmidFinder web-service (http://cge.cbs.dtu.dk/services/PlasmidFinder) was used to identify replicons in the assembled contigs and classify plasmids into incompatibility groups [37]. Plasmid ring diagrams depicting read coverage were drawn in R via custom scripts [54], and plasmid ring diagrams were drawn using BLAST Ring Image Generator (BRIG) [55] with the ?-task megablast? option to BLAST [56]. Additionally, contigs belonging to plasmids that have a copy number greater than one were also identified based on their relative abundance to the genome. Read depth was calculated via BEDTools [26]. Plasmid sequence comparisons for pNK117-2 were made to pOLA52 (NC_010378.1), pRPEC180_47 (NC_019088.1) and NDM-1 producing plasmid from RJA274 (KF877335.1). IncF pMLST typing was done using the web-service (https://cge.cbs.dtu.dk/services/pMLST/) [37].

Genomic Deletion Verification

Based on the alignment of contigs to the genome of CFT 073 (NC_004431), primers were designed to show that the deletion in lineage A was an excision, in addition to show contiguity prior to the deletion, as well as controls for showing the occurrence of the deletion only in the lineage A isolate sampled at 6 months.

Competition experiments to assess the fitness costs in vitro of carriage of the plasmid closely resembling pHUSEC41-1

To assess fitness cost, pair-wise growth competition in Davis minimal medium with 25 mg/mL glucose (DM25) was performed using isolates of lineage A sampled at 2 weeks and 2 months, respectively, the latter which had acquired the plasmid closely resembling pHUSEC41-1. The experiment was performed as previously described [57], but in brief, the two isolates were grown overnight in nutrient both, and then inoculated into DM25 at a dilution of $1{:}10^4$ and grown for 24 hours. The cultures were then mixed together in a ratio of 1:1, and then diluted 1:100 into fresh DM25. This dilution step was continued for 6 days. After initially mixing the two cultures together, and after each 24 hour period, the cultures were diluted appropriately and $100u$L was added to Iso-Sensitest plates (Oxoid, Sweden) in triplicate, with and without 1000 mg/L of sulfamethoxazole. Colonies were counted after over night incubation at 37°C, and the mean number of colonies on the sulfamethoxazole plates were subtracted

from the plates without sulfamethoxazole to determine the mean number of colonies lacking the pHUSEC41-1-like plasmids. Six replicates of this fitness experiment were conducted.

4.6 Acknowledgements

We would like to thank Mari Rodriguez de Evgrafov for preparing single-end sequencing libraries, Andreas Porse for preparing the paired-end sequencing libraries and Lejla Imamovic for advice regarding phages. This work was supported by the Danish Free Research Councils for Health and Disease, the European Union FP7-HEALTH-2011-single-stage grant agreement 282004, EvoTAR, the Medical Faculty of the University of Göteborg (ALFGBG138401) and the Swedish Medical Research Council. M.O.A.S further acknowledges support from the Novo Nordisk Foundation, and the Lundbeck Foundation

Author Contributions

NK, IA and AW designed the study, made the original finding and analyzed the preliminary data regarding transfer of a resistance phenotype and a conjugative resistance plasmid between E. coli lineages in situ. NK performed all experiments in the early stage of the study. HG processed the sequencing data and performed experiments in the late stage of the study. HG, CM and MS analyzed all sequencing data and wrote the manuscript with input from NK, IA and AW.

4.7 References

[1] J Davies and D Davies. Origins and Evolution of Antibiotic Resistance. *Microbiology and Molecular Biology Reviews*, 74(3):417–433, August 2010.

[2] Lei Yang, Lars Jelsbak, Rasmus Lykke Marvig, Søren Damkiær, Christopher T Workman, Martin Holm Rau, Susse Kirkelund Hansen, Anders Folkesson, Helle Krogh Johansen, Oana Ciofu, Niels Høiby, Morten O A Sommer, and Søren Molin. Evolutionary dynamics of bacteria in a human host environment. *Proceedings of the National Academy of Sciences*, 108(18):7481–7486, May 2011.

[3] P M Hawkey and A M Jones. The changing epidemiology of resistance. *Journal of Antimicrobial Chemotherapy*, 64(Supplement 1):i3–i10, August 2009.

[4] Vanessa M D'Costa, Katherine M McGrann, Donald W Hughes, and Gerard D Wright. Sampling the antibiotic resistome. *Science (New York, N.Y.)*, 311(5759):374–377, January 2006.

[5] K J Forsberg, A Reyes, B Wang, E M Selleck, M O A Sommer, and G Dantas. The Shared Antibiotic Resistome of Soil Bacteria and Human Pathogens. *Science (New York, N.Y.)*, 337(6098):1107–1111, August 2012.

[6] M O A Sommer, G Dantas, and G M Church. Functional Characterization of the Antibiotic Resistance Reservoir in the Human Microflora. *Science (New York, N.Y.)*, 325(5944):1128–1131, August 2009.

[7] Morten OA Sommer and Gautam Dantas. Antibiotics and the resistant microbiome. *Current Opinion in Microbiology*, 14(5):556–563, October 2011.

[8] Abigail A Salyers, Anamika Gupta, and Yanping Wang. Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends in microbiology*, 12(9):412–416, September 2004.

[9] K Forslund, S Sunagawa, J Roat Kultima, D Mende, M Arumugam, A Typas, and P Bork. Country-specific antibiotic use practices impact the human gut resistome. *Genome Research*, April 2013.

[10] Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402):207–214, April 2013.

[11] Junjie Qin, Ruiqiang Li, Jeroen Raes, Manimozhiyan Arumugam, Kristoffer Solvsten Burgdorf, Chaysavanh Manichanh, Trine Nielsen, Nicolas Pons, Florence Levenez, Takuji Yamada, Daniel R Mende, Junhua Li, Junming Xu, Shaochuan Li, Dongfang Li, Jianjun Cao, Bo Wang, Huiqing Liang, Huisong Zheng, Yinlong Xie, Julien Tap, Patricia Lepage, Marcelo Bertalan, Jean-Michel Batto, Torben Hansen, Denis Le Paslier, Allan Linneberg, H Bjørn Nielsen, Eric Pelletier, Pierre Renault, Thomas Sicheritz-Ponten, Keith Turner, Hongmei Zhu, Chang Yu, Shengting Li, Min Jian, Yan Zhou, Yingrui Li, Xiuqing Zhang, Songgang Li, Nan Qin, Huanming Yang, Jian Wang, Søren Brunak, Joel Dore, Francisco Guarner, Karsten Kristiansen, Oluf Pedersen, Julian Parkhill, Jean Weissenbach, MetaHIT Consortium, Peer Bork, and S Dusko Ehrlich Jun Wang. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285):59–65, April 2010.

[12] I Sharon, M J Morowitz, B C Thomas, E K Costello, D A Relman, and J F Banfield. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Research*, 23(1):111–120, January 2013.

[13] M A McConnell, A A Mercer, and G W Tannock. Transfer of Plasmid pAMβl Between Members of the Normal Microflora Inhabiting the Murine Digestive Tract and Modification of the Plasmid in a Lactobacillus reuteri Host. *Microbial ecology in health and disease*, 4(6):343–355, 1991.

[14] C H Lester, N Frimodt-Moller, T L Sorensen, D L Monnet, and A M Hammerum. In Vivo Transfer of the vanA Resistance Gene from an Enterococcus faecium Isolate of Animal Origin to an E. faecium Isolate of Human Origin in the Intestines of Human Volunteers. *Antimicrobial agents and chemotherapy*, 50(2):596–599, January 2006.

[15] M Trobos, C H Lester, J E Olsen, N Frimodt-Moller, and A M Hammerum. Natural transfer of sulphonamide and ampicillin resistance between Escherichia coli residing in the human intestine. *Journal of Antimicrobial Chemotherapy*, 63(1):80–86, October 2008.

[16] Bärbel Stecher, Rémy Denzler, Lisa Maier, Florian Bernet, Mandy J Sanders, Derek J Pickard, Manja Barthel, Astrid M Westendorf, Karen A Krogfelt, and Alan W Walker. Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proceedings of the National Academy of Sciences*, 109(4):1269–1274, 2012.

[17] P Bidet, B Burghoffer, V Gautier, N Brahimi, P Mariani-Kurkdjian, A El-Ghoneimi, E Bingen, and G Arlet. In Vivo Transfer of Plasmid-Encoded ACC-1 AmpC from Klebsiella pneumoniae to Escherichia coli in an Infant and Selection of Impermeability to Imipenem in K. pneumoniae. *Antimicrobial agents and chemotherapy*, 49(8):3562–3565, July 2005.

[18] N Karami, A Martner, V I Enne, S Swerkersson, I Adlerberth, and A E Wold. Transfer of an ampicillin resistance gene between two Escherichia coli strains in the bowel microbiota of an infant treated with antibiotics. *Journal of Antimicrobial Chemotherapy*, 60(5):1142–1145, September 2007.

[19] N B Shoemaker, H Vlamakis, K Hayes, and A A Salyers. Evidence for Extensive Resistance Gene Transfer among Bacteroides spp. and among Bacteroides and Other Genera in the Human Colon. *Applied and Environmental Microbiology*, 67(2):561–568, February 2001.

[20] Forough Nowrouzian, Bill Hesselmar, Robert Saalman, Inga-Lisa Strannegård, Nils Åberg, Agnes E Wold, and Ingegerd Adlerberth. Escherichia coli in Infants' Intestinal Microflora: Colonization Rate, Strain Turnover, and Virulence Gene Carriage. *Pediatric Research*, 54(1):8–14, July 2003.

[21] Forough L Nowrouzian, Agnes E Wold, and Ingegerd Adlerberth. Escherichia coli strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *The Journal of infectious diseases*, 191(7):1078–1083, April 2005.

[22] D R Zerbino and E Birney. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18(5):821–829, February 2008.

[23] Ramy K Aziz, Daniela Bartels, Aaron A Best, Matthew DeJongh, Terrence Disz, Robert A Edwards, Kevin Formsma, Svetlana Gerdes, Elizabeth M Glass, Michael Kubal, Folker Meyer, Gary J Olsen, Robert Olson, Andrei L Osterman, Ross A Overbeek, Leslie K McNeil, Daniel Paarmann, Tobias Paczian, Bruce Parrello, Gordon D Pusch, Claudia Reich, Rick Stevens, Olga Vassieva, Veronika Vonstein, Andreas Wilke, and Olga Zagnitko. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC genomics*, 9(1):75, 2008.

[24] Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4):357–359, April 2012.

[25] H Li, B Handsaker, A Wysoker, T Fennell, J Ruan, N Homer, G Marth, G Abecasis, R Durbin, and 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, August 2009.

[26] A R Quinlan and I M Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, March 2010.

[27] Stefan Kurtz, Adam Phillippy, Arthur L Delcher, Michael Smoot, Martin Shumway, Corina Antonescu, and Steven L Salzberg. Versatile and open software for comparing large genomes. *Genome Biology*, 5(2):R12, 2004.

[28] Jaroslaw Zdziarski, Elzbieta Brzuszkiewicz, Björn Wullt, Heiko Liesegang, Dvora Biran, Birgit Voigt, Jenny Grönberg-Hernandez, Bryndis Ragnarsdottir, Michael Hecker, Eliora Z Ron, Rolf Daniel, Gerhard Gottschalk, Jörg Hacker, Catharina Svanborg, and Ulrich Dobrindt. Host Imprints on Bacterial Genomes—Rapid, Divergent Evolution in Individual Patients. *PLoS pathogens*, 6(8):e1001078, August 2010.

[29] Marie Touchon, Claire Hoede, Olivier Tenaillon, Valérie Barbe, Simon Baeriswyl, Philippe Bidet, Edouard Bingen, Stéphane Bonacorsi, Christiane Bouchier, Odile Bouvet, Alexandra Calteau, Hélène Chiapello, Olivier Clermont, Stéphane Cruveiller, Antoine Danchin, Médéric Diard, Carole Dossat, Meriem El Karoui, Eric Frapy, Louis Garry, Jean Marc Ghigo, Anne Marie Gilles, James Johnson, Chantal Le Bouguénec, Mathilde Lescat, Sophie Mangenot, Vanessa Martinez-Jéhanne, Ivan Matic, Xavier Nassif, Sophie Oztas, Marie Agnès Petit, Christophe Pichon, Zoé Rouy, Claude Saint Ruf, Dominique Schneider, Jérôme Tourret, Benoit Vacherie, David Vallenet, Claudine Médigue, Eduardo P C Rocha, and Erick Denamur. Organised Genome Dynamics in the Escherichia coli Species Results in Highly Diverse Adaptive Paths. *PLoS Genetics*, 5(1):e1000344, January 2009.

[30] M Lescat, A Calteau, C Hoede, V Barbe, M Touchon, E Rocha, O Tenaillon, C Medigue, J R Johnson, and E Denamur. A Module Located at a Chromosomal Integration Hot Spot Is Responsible for the Multidrug Resistance of a Reference Strain from Escherichia coli Clonal Group A. *Antimicrobial agents and chemotherapy*, 53(6):2283–2288, May 2009.

[31] J R Johnson, K Owens, A R Manges, and L W Riley. Rapid and Specific Detection of Escherichia coli Clonal Group A by Gene-Specific PCR. *Journal of clinical microbiology*, 42(6):2618–2622, June 2004.

[32] D M Livermore. beta-Lactamases in laboratory and clinical resistance. *Clinical microbiology reviews*, 8(4):557–584, October 1995.

[33] Maria S Ramirez and Marcelo E Tolmasky. Drug Resistance Updates. *Drug Resistance Updates*, 13(6):151–171, December 2010.

[34] Ola Sköld. Sulfonamide resistance: mechanisms and trends. *Drug Resistance Updates*, 3(3):155–160, June 2000.

[35] C Kunne, A Billion, S E Mshana, J Schmiedel, E Domann, H Hossain, T Hain, C Imirzalioglu, and T Chakraborty. Complete Sequences of Plasmids from the Hemolytic-Uremic Syndrome-Associated Escherichia coli Strain HUSEC41. *Journal of bacteriology*, 194(2):532–533, December 2011.

[36] Yonatan H Grad, Paul Godfrey, Gustavo C Cerquiera, Patricia Mariani-Kurkdjian, Malika Gouali, Edouard Bingen, Terrence P Shea, Brian J Haas, Allison Griggs, Sarah Young, Qiandong Zeng, Marc Lipsitch, Matthew K Waldor, François-Xavier Weill, Jennifer R Wortman, and William P Hanage. Comparative genomics of recent Shiga toxin-producing Escherichia coli O104:H4: short-term evolution of an emerging pathogen. *mBio*, 4(1):e00452–12, 2013.

[37] A Carattoli, E Zankari, A Garcia-Fernandez, M Voldby Larsen, O Lund, L Villa, F Moller Aarestrup, and H Hasman. In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrobial agents and chemotherapy*, 58(7):3895–3903, June 2014.

[38] L Villa, A Garcia-Fernandez, D Fortini, and A Carattoli. Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. *Journal of Antimicrobial Chemotherapy*, 65(12):2518–2529, November 2010.

[39] Joel Selkrig, Khedidja Mosbahi, Chaille T Webb, Matthew J Belousoff, Andrew J Perry, Timothy J Wells, Faye Morris, Denisse L Leyton, Makrina Totsika, Minh-Duy Phan, Nermin Celik, Michelle Kelly, Clare Oates, Elizabeth L Hartland, Roy M Robins-Browne, Sri Harsha Ramarathinam, Anthony W Purcell, Mark A Schembri, Richard A Strugnell, Ian R Henderson, Daniel Walker, and Trevor Lithgow. Discovery of an archetypal protein transport system in bacterial outer membranes. *Nature Structural & Molecular Biology*, 19(5):506–510, April 2012.

[40] M C Lane and H L T Mobley. Role of P-fimbrial-mediated adherence in pyelonephritis and persistence of uropathogenic Escherichia coli (UPEC) in the mammalian kidney. *Kidney International*, 72(1):19–25, March 2007.

[41] F Nowrouzian, I Adlerberth, and A E Wold. P fimbriae, capsule and aerobactin characterize colonic resident Escherichia coli. *Epidemiology and infection*, 126(1):11–18, February 2001.

[42] Forough L Nowrouzian, Ingegerd Adlerberth, and Agnes E Wold. Enhanced persistence in the colonic microbiota of Escherichia coli strains belonging to phyloge-

netic group B2: role of virulence factors and adherence to colonic cells. *Microbes and Infection*, 8(3):834–840, March 2006.

[43] Timothy J Johnson, Eliza M Bielak, Daniela Fortini, Lars Hestbjerg Hansen, Henrik Hasman, Chitrita Debroy, Lisa K Nolan, and Alessandra Carattoli. Expansion of the IncX plasmid family for improved identification and typing of novel plasmids in drug-resistant Enterobacteriaceae. *Plasmid*, 68(1):43–50, July 2012.

[44] Swaine L Chen, Chia-Seui Hung, Jian Xu, Christopher S Reigstad, Vincent Magrini, Aniko Sabo, Darin Blasiar, Tamberlyn Bieri, Rekha R Meyer, and Philip Ozersky. Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a comparative genomics approach. *Proceedings of the National Academy of Sciences of the United States of America*, 103(15):5977–5982, 2006.

[45] D O Krause, A C Little, S E Dowd, and C N Bernstein. Complete Genome Sequence of Adherent Invasive Escherichia coli UM146 Isolated from Ileal Crohn's Disease Biopsy Tissue. *Journal of bacteriology*, 193(2):583–583, December 2010.

[46] Chitrita Debroy, Mandeep S Sidhu, Upal Sarker, Bhushan M Jayarao, Adam L Stell, Nathan P Bell, and Timothy J Johnson. Complete sequence of pEC14_114, a highly conserved IncFIB/FIIA plasmid associated with uropathogenic Escherichia coli cystitis strains. *Plasmid*, 63(1):53–60, January 2010.

[47] Dona Wijetunge, Kurundu Karunathilake, Atul Chaudhari, Robab Katani, Edward G Dudley, Vivek Kapur, Chitrita Debroy, and Subhashinie Kariyawasam. Complete nucleotide sequence of pRS218, a large virulence plasmid that augments pathogenic potential of meningitis-associated. *BMC Microbiology*, 14(1):203, August 2014.

[48] Fritz Kauffmann. *The bacteriology of enterobacteriaceae*. Munksgaard, Copenhagen, 1969.

[49] Olivier Clermont, Stéphane Bonacorsi, and Edouard Bingen. Rapid and Simple Determination of theEscherichia coli Phylogenetic Group. *Applied and Environmental Microbiology*, 66(10):4555–4558, 2000.

[50] P J A Cock, T Antao, J T Chang, B A Chapman, C J Cox, A Dalke, I Friedberg, T Hamelryck, F Kauff, B Wilczynski, and M J L de Hoon. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423, May 2009.

[51] Z Khan, J S Bloom, L Kruglyak, and M Singh. A practical algorithm for finding maximal exact matches in large sequence datasets using sparse suffix arrays. *Bioinformatics*, 25(13):1609–1616, June 2009.

[52] Thierry Wirth, Daniel Falush, Ruiting Lan, Frances Colles, Patience Mensa, Lothar H Wieler, Helge Karch, Peter R Reeves, Martin C J Maiden, Howard Ochman, and Mark Achtman. Sex and virulence in Escherichia coli: an evolutionary perspective. *Molecular microbiology*, 60(5):1136–1151, June 2006.

[53] Y Zhou, Y Liang, K H Lynch, J J Dennis, and D S Wishart. PHAST: A Fast Phage Search Tool. *Nucleic Acids Research*, 39(suppl):W347–W352, June 2011.

[54] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, October 2012.

[55] Nabil-Fareed Alikhan, Nicola K Petty, Nouri L Ben Zakour, and Scott A Beatson. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC genomics*, 12(1):402, August 2011.

[56] S F Altschul, W Gish, W Miller, E W Myers, and D J Lipman. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410, October 1990.

[57] Virve I Enne, Peter M Bennett, David M Livermore, and Lucinda M C Hall. Enhancement of host fitness by the sul2-coding plasmid p9123 in the absence of selective pressure. *The Journal of antimicrobial chemotherapy*, 53(6):958–963, June 2004.

4.8 Supplementary Information

Supplementary Table 4.1| MIC values of lineages A, B and C for sampling time points. One isolate per lineage per sampling time was selected for MIC testing.

Lineage	Lineage A				Lineage B			Lineage C
Sampling Time	$2w_1$	4w	2m	$6m_1$	$2m_1$	6m	$12m_1$	$12m_1$
Antibiotic (ug/mL)								
Ampicillin	2	2	512	512	512	512	512	2
Piperacillin	1	1	64	64	32	32	32	0.5
Mecillinam	0.125	0.125	1	1	1	1	0.5	0.125
Ceftazidime	0.25	0.25	0.25	0.25	0.0625	0.125	0.125	0.125
Cefuroxime	8	8	8	8	2	4	2	2
Cefoxitin	2	4	4	4	8	4	8	2
Chloramphenicol	1	1	1	1	1	1	1	1
Gentamycin	0.5	0.5	0.5	0.5	1	1	1	1
Tobramycin	0.25	0.25	0.5	0.125	0.5	0.25	0.5	0.5
Streptomycin	2	2	64	128	128	128	128	4
Nalidixic Acid	1	1	1	1	1	0.5	1	0.5
Tetracycline	0.5	0.5	0.5	0.5	1	1	0.5	1
Trimethoprim	4	4	2	4	0.125	0	0.125	0.125
Sulfamethoxazole	8	16	1024	1024	512	1024	1024	512

Supplementary Table 4.2| Overview of library preparation method and sequencing technology used for the isolates genome sequenced in this study.

Lineage	Sampling Time	Library Preparation	Sequencing Technology Output
A	$2w_1$	Shear+barcode	SE
	$2w_2$	TruSeq kit	PE
	4w	Shear+barcode	SE
	2m	Shear+barcode	SE
	$6m_1$	Shear+barcode	SE
	$6m_2$	TruSeq kit	PE
B	$2m_1$	Shear+barcode	SE
	$2m_2$	TruSeq kit	PE
	6m	Shear+barcode	SE
	$12m_1$	Shear+barcode	SE
	$12m_2$	TruSeq kit	PE
C	$12m_1$	Shear+barcode	SE
	$12m_2$	TruSeq kit	PE

Supplementary Table 4.3| Number of SNPs between the lineages using selected isolates. The rows of the table indicate the reads that were used aligned to contigs of the isolate as indicated in the column.

		Aligned to contigs from		
		Lineage A $2w_2$	Lineage B $2m_2$	Lineage C $12m_2$
Reads from	Lineage A $2w_2$		94346	7324
	Lineage B $2m_2$	95215		94797
	Lineage C $12m_2$	7297	92100	

Supplementary Table 4.4| Coverage of the total genomic content between the lineages using selected isolates. The rows of the table indicate the reads that were used aligned to contigs of the isolate as indicated in the column.

		Aligned to contigs from		
		Lineage A $2w_2$	Lineage B $2m_2$	Lineage C $12m_2$
Reads from	A $2w_2$		80.0%	93.60%
	B $2m_2$	82.3%		82.0%
	C $12m_2$	95.5%	79.7%	

Supplementary Table 4.5| Table containing the SNPs from lineage B, including the annotation and whether the animo acid change was synonymous or nonsynonymous.

Contig number	Position	$2w_2$	$2w_1$	4w	2m	$6m_1$	$6m_2$	Annotation	Change
5	27035	A		G	G	G	G		
7	282598	G			A			Formate hydrogenlyase subunit 2	synonymous
32	38337	A			C			Betaine aldehyde dehydrogenase (EC 1.2.1.8)	nonsynonymous
13	108313	C			T			Phosphoenolpyruvate carboxylase (EC 4.1.1.31)	nonsynonymous
12	157912	G				A		Putative inner membrane protein	synonymous
52	28325	C				A		Protein fdrA	nonsynonymous
239	107	T				C			
93	2581	T					C		
8	11719	G					T	type 1 fimbriae anchoring protein FimD	nonsynonymous

Supplementary Table 4.6| Table containing the SNPs from lineage B, including the annotation and whether the animo acid change was synonymous or nonsynonymous.

Contig number	Position	$2m_2$	$2m_1$	6m	$12m_1$	$12m_2$	Annotation	Change
12	273551	T		A			Putative amidohydrolase	nonsynonymous
20	13798	T		A			FIG00639301: hypothetical protein	nonsynonymous
590	436	G		A			Phage major tail tube protein	synonymous
91	20935	G		A			Alcohol dehydrogenase (EC 1.1.1.1)	nonsynonymous
709	463	G		A				
141	121	T		C				
64	121	T		C				
237	2376	C		G			IncF plasmid conjugative transfer pilin protein TraA	nonsynonymous
37	40257	A		G			LysR family transcriptional regulator lrhA	synonymous
389	490	T		G			Phage capsid scaffolding protein	nonsynonymous
56	41072	T		G			Dipeptide transport system permease protein DppB (TC 3.A.1.5.2)	nonsynonymous
58	4945	A		G				
24	63963	A		T	T	T	Oligopeptide transport system permease protein OppB (TC 3.A.1.5.1)	nonsynonymous
28	44283	C		T			Fe2+-dicitrate sensor, membrane component	nonsynonymous
35	41468	A		T			C-terminal domain of CinA type S	nonsynonymous
511	121	C		T			Adenine-specific methyltransferase (EC 2.1.1.72)	nonsynonymous
63	19848	A		T			Type III secretion inner membrane protein (YscU,SpaS,EscU,HrcU,SsaU, homologous to flagellar export components)	nonsynonymous
593	514	C		T				
778	418	G		T				
36	2740	T			A	A	Chaperone protein DnaK	nonsynonymous
37	78302	T			A	A	Putative transporting ATPase	nonsynonymous
50	33170	C			A		Putative vimentin	nonsynonymous
80	12348	G			A		D-Galactonate repressor DgoR	nonsynonymous
12	327228	G			A			
42	91228	G			C	C	FUSARIC ACID RESISTANCE PROTEIN FUSB / FUSARIC ACID RESISTANCE PROTEIN FUSC	nonsynonymous
18	98805	A			T	T	N-succinyl-L,L-diaminopimelate desuccinylase (EC 3.5.1.18)	nonsynonymous

104	5105	C	T		LSU ribosomal protein L22p (L17e)	nonsynonymous
8	149794	A	T			
7	156	T		A	hypothetical protein	synonymous
776	244	T		A		
7	121	A		C	hypothetical protein	nonsynonymous
80	12641	T		G		
12	191888	C		T	NADP-specific glutamate dehydrogenase (EC 1.4.1.4)	synonymous
121	10485	C		T	Oligopeptide transport ATP-binding protein OppD (TC 3.A.1.5.1)	nonsynonymous

Supplementary Table 4.7| Table containing the SNPs from lineage C, including the annotation and whether the animo acid change was synonymous or nonsynonymous.

Contig number	Position	$12m_1$	$12m_2$	Annotation	Change
35	132	T	C		

Supplementary Table 4.8| Multi locus sequence typing (MLST) of the lineages.

Lineage	adk	fumC	gyrB	icd	mdh	purA	recA	Sequence Type
A	13	13	9	13	16	10	9	ST12
B	21	35	27	6	5	5	99	ST782
C	13	13	9	13	16	10	9	ST12

Supplementary Table 4.9| Replicon type results from PlasmidFinder, presence in the isolates and the corresponding plasmid. Subvariants of the replicon type from PlasmidFinder are given in parentheses.

Lineage	Isolates	Identified Replicon	% Identity	Associated Plasmid
A	2m, $6m_{1,2}$	B/O/K/Z	100	pHUSEC41-1
B	all	B/O/K/Z	100	pHUSEC41-1
B	$2m_{1,2}$,$12m_{1,2}$	FIB(AP001918)	99.56	
B	$2m_{1,2}$, 6m	FII	96.55	
B	$2m_{1,2}$,$12m_{1,2}$	FII	92.34	
B	$2m_{1,2}$,$12m_{1,2}$	FIC	96.2	
B	all	Col156	94.81	
B	all	Col8282	89.9	
B	all	ColpVC	88.95	
C	all	none identified		

Supplementary Table 4.10| Annotations of the FII plasmid from the 6m isolate done by RAST. The FII plasmid was observed in 2m and 6m isolates.

Contig number	Start	Stop	Strand	Annotation
48	72	359	+	FIG00638017: hypothetical protein
48	478	1299	+	unnamed protein product
48	1595	2242	-	X polypeptide
48	2519	2902	+	IncF plasmid conjugative transfer mating signal transduction protein TraM
48	3093	3779	+	IncF plasmid conjugative transfer regulator TraJ
48	3918	4100	+	IncF plasmid conjugative transfer regulator TraY
48	4134	4493	+	IncF plasmid conjugative transfer pilin protein TraA
48	4508	4819	+	IncF plasmid conjugative transfer pilus assembly protein TraL
48	4841	5407	+	IncF plasmid conjugative transfer pilus assembly protein TraE
48	5394	6122	+	IncF plasmid conjugative transfer pilus assembly protein TraK
48	6122	7549	+	IncF plasmid conjugative transfer pilus assembly protein TraB
48	7539	8129	+	IncF plasmid conjugative transfer protein TraP
48	8116	8436	+	IncF plasmid conjugative transfer protein TrbD
48	8429	8680	+	IncF plasmid conjugative transfer protein TrbG
48	8677	9192	+	IncF plasmid conjugative transfer pilus assembly protein TraV
48	9327	9548	+	IncF plasmid conjugative transfer protein TraR
48	9708	12335	+	IncF plasmid conjugative transfer pilus assembly protein TraC
48	12332	12718	+	IncF plasmid conjugative transfer protein TrbI
48	12715	13347	+	IncF plasmid conjugative transfer pilus assembly protein TraW
48	13344	14336	+	IncF plasmid conjugative transfer pilus assembly protein TraU
48	14362	14823	+	hypothetical protein
48	15054	15305	+	hypothetical protein
48	15333	15971	+	IncF plasmid conjugative transfer protein TrbC
48	15968	17776	+	IncF plasmid conjugative transfer protein TraN
48	17804	18061	+	IncF plasmid conjugative transfer protein TrbE
48	18054	18797	+	IncF plasmid conjugative transfer pilus assembly protein TraF
48	18813	19160	+	IncF plasmid conjugative transfer protein TrbA

Contig number	Start	Stop	Strand	Annotation
48	19162	19476	-	Protein artA
48	19355	19492	+	FIG00640935: hypothetical protein
48	19557	19841	+	IncF plasmid conjugative transfer protein TraQ
48	19828	20373	+	IncF plasmid conjugative transfer protein TrbB
48	20363	20644	+	IncF plasmid conjugative transfer protein TrbJ
48	20631	21011	+	IncF plasmid conjugative transfer protein TrbF
48	21011	22384	+	IncF plasmid conjugative transfer pilus assembly protein TraH
48	22381	25203	+	IncF plasmid conjugative transfer protein TraG
48	25748	26479	+	IncF plasmid conjugative transfer surface exclusion protein TraT
48	26731	28884	+	IncF plasmid conjugative transfer protein TraD
48	28893	29291	-	VapC toxin protein
48	29291	29518	-	VapB protein (antitoxin to VapC)
48	29672	34870	+	IncF plasmid conjugative transfer DNA-nicking and unwinding protein TraI
48	34890	35636	+	IncF plasmid conjugative transfer pilin acetylase TraX
48	35695	36555	+	Dienelactone hydrolase and related enzymes
48	36658	37218	+	IncF plasmid conjugative transfer fertility inhibition protein FinO
48	37351	37509	+	FIG00640597: hypothetical protein
48	37481	37606	+	FIG00639321: hypothetical protein
48	37808	38269	+	FIG01068181: hypothetical protein
48	38389	38535	-	FIG00640641: hypothetical protein
48	38562	39152	+	YihA
48	39392	39646	+	Replication regulatory protein repA2 (Protein copB)
48	39949	40806	+	RepA1
712	1446	2636	-	microcin b17 processing protein mcbd, putative
712	2617	3435	-	hypothetical protein
712	3437	4291	-	Microcin B17-processing protein mcbB
712	4338	4463	-	hypothetical protein
712	5004	5300	+	COG1266: Predicted metal-dependent membrane protease
712	5345	5797	+	Acetyltransferase (EC 2.3.1.-);Ontology_term=KEGG_ENZYME:2.3.1.-
712	6553	7515	-	Putative stability/partitioning protein encoded within prophage CP-933T
712	7799	7924	-	hypothetical protein
501	761	1426	-	Fip
501	1625	1744	+	hypothetical protein
340	734	1048	-	FIG00641039: hypothetical protein
340	1045	1764	-	Plasmid SOS inhibition protein PsiA
340	1761	2159	-	PsiB protein

Supplementary Table 4.11| Annotations of the FII:FIB:FIC plasmid from the $12m_1$ isolate done by RAST. The FII:FIB:FIC plasmid was observed in 2m and 12m isolates.

Contig number	Start	Stop	Strand	Annotation
65	19	819	-	Mobile element protein
65	2144	3001	-	RepA1
65	3300	3557	-	replication protein
65	4382	4645	-	FIG00639321: hypothetical protein
65	4780	5340	-	IncF plasmid conjugative transfer fertility inhibition protein FinO
65	5395	6135	-	IncF plasmid conjugative transfer pilin acetylase TraX
65	6155	11353	-	IncF plasmid conjugative transfer DNA-nicking and unwinding protein TraI
65	11425	13605	-	IncF plasmid conjugative transfer protein TraD
65	13858	14589	-	IncF plasmid conjugative transfer surface exclusion protein TraT
65	14638	14802	-	IncF plasmid conjugative transfer surface exclusion protein TraS
65	15139	17982	-	IncF plasmid conjugative transfer protein TraG
65	17979	19355	-	putative conjugative transfer protein TraH
65	19352	19654	-	IncF plasmid conjugative transfer protein TrbJ
65	19644	20189	-	IncF plasmid conjugative transfer protein TrbB
65	20176	20460	-	IncF plasmid conjugative transfer protein TraQ
65	20579	20926	-	IncF plasmid conjugative transfer protein TrbA
65	20942	21685	-	IncF plasmid conjugative transfer pilus assembly protein TraF
65	21678	21935	-	IncF plasmid conjugative transfer protein TrbE
65	21962	23770	-	IncF plasmid conjugative transfer protein TraN
65	23767	24405	-	IncF plasmid conjugative transfer protein TrbC
65	24433	24885	-	hypothetical protein
65	24915	25376	-	hypothetical protein
65	25402	26394	-	IncF plasmid conjugative transfer pilus assembly protein TraU
65	26391	27023	-	IncF plasmid conjugative transfer pilus assembly protein TraW
65	27020	27406	-	IncF plasmid conjugative transfer protein TrbI
65	27403	30030	-	IncF plasmid conjugative transfer pilus assembly protein TraC
65	30190	30411	-	IncF plasmid conjugative transfer protein TraR
65	30545	31060	-	IncF plasmid conjugative transfer pilus assembly protein TraV
65	31057	31308	-	IncF plasmid conjugative transfer protein TrbG
65	31305	31622	-	IncF plasmid conjugative transfer protein TrbD
65	31619	32191	-	IncF plasmid conjugative transfer protein TraP
65	32181	33608	-	IncF plasmid conjugative transfer pilus assembly protein TraB
65	33608	34336	-	IncF plasmid conjugative transfer pilus assembly protein TraK
65	34323	34889	-	IncF plasmid conjugative transfer pilus assembly protein TraE
65	34911	35222	-	IncF plasmid conjugative transfer pilus assembly protein TraL
65	35237	35602	-	IncF plasmid conjugative transfer pilin protein TraA
65	36831	37214	-	IncF plasmid conjugative transfer mating signal transduction protein TraM
65	37582	38139	+	X polypeptide
65	38436	39257	-	unnamed protein product
65	39376	39663	-	FIG00638017: hypothetical protein
90	157	897	+	hypothetical protein
90	1056	1193	-	hypothetical protein
90	1186	1998	-	FIG01047974: hypothetical protein
90	2784	3761	+	RepFIB replication protein A
90	3837	3956	-	hypothetical protein
90	4040	4780	-	Resolvase
90	4901	5026	-	hypothetical protein

Contig number	Start	Stop	Strand	Annotation
90	5112	5237	-	FIG00643146: hypothetical protein
90	5429	5602	-	hypothetical protein
90	6392	6826	-	probable copper-binding protein
90	6825	6962	+	hypothetical protein
90	7044	8444	-	Heavy metal sensor histidine kinase
90	8441	9121	-	DNA-binding heavy metal response regulator
90	9176	9919	-	Copper resistance protein D
90	9998	10123	+	hypothetical protein
90	10110	10448	-	Copper resistance protein CopC
90	10530	11420	-	Copper resistance protein B
90	11426	13243	-	Multicopper oxidase
90	13478	13927	+	Copper-binding protein PcoE
90	14216	14953	+	Cell wall endopeptidase, family M23/M37
90	14987	15184	-	FIG00643033: hypothetical protein
90	15225	17030	-	Lead, cadmium, zinc and mercury transporting ATPase (EC 3.6.3.3) (EC 3.6.3.5); Copper-translocating P-type ATPase (EC 3.6.3.4);Ontology_term
90	17145	17615	+	hypothetical protein
90	17846	18286	-	CopG protein
90	18372	21518	-	Cobalt-zinc-cadmium resistance protein CzcA; Cation efflux system protein CusA
90	21529	22821	-	Cobalt/zinc/cadmium efflux RND transporter, membrane fusion protein, CzcB family
90	22935	23288	-	Cation efflux system protein CusF precursor
90	23317	24702	-	Cation efflux system protein CusC precursor
90	24668	24832	+	hypothetical protein
90	24892	25572	+	Copper-sensing two-component system response regulator CusR
90	25565	27040	+	Osmosensitive K+ channel histidine kinase KdpD (EC 2.7.3.-);Ontology_term
90	27291	27722	+	Silver-binding protein
90	28218	29048	-	Mobile element protein
90	29075	29338	-	Mobile element protein
90	29935	30114	+	hypothetical protein
90	30108	30278	-	hypothetical protein
90	31003	32025	-	Mobile element protein
90	32248	32853	+	TnpA transposase
127	141	845	-	Mobile element protein
127	836	1342	+	putative protein
127	1339	2130	+	Resolvase
127	2268	2543	-	FIG00641502: hypothetical protein
127	2537	2800	-	Plasmid partitioning protein ParA
127	3410	4381	+	Putative stability/partitioning protein encoded within prophage CP-933T
127	4386	4778	+	stable plasmid inheritance protein
203	213	1040	-	UspA-related nucleotide-binding protein
203	1059	2537	-	Sulfate permease
203	2925	3509	-	DNA-invertase
203	3669	6653	+	Mobile element protein
215	318	1721	+	S-methylmethionine permease
215	1708	2640	+	Homocysteine S-methyltransferase (EC 2.1.1.10);Ontology_term=KEGG_ENZYME:2.1.1.10
215	2749	3795	-	ABC transporter ATP-binding protein
215	3807	4169	-	Iron(III)-transport system permease protein SfuB

Contig number	Start	Stop	Strand	Annotation
238	247	3363	-	Type I restriction-modification system, restriction subunit R (EC 3.1.21.3);Ontology_term=KEGG_ENZYME:3.1.21.3
238	3485	4711	-	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3);Ontology_term=KEGG_ENZYME:3.1.21.3
238	4708	6264	-	Type I restriction-modification systemm DNA-methyltransferase subunit M (EC 2.1.1.72);Ontology_term=KEGG_ENZYME:2.1.1.72
238	6447	6668	+	Prevent host death protein, Phd antitoxin
238	6668	7048	+	Death on curing protein, Doc toxin
238	7053	7232	+	PdcA
238	7260	7619	+	FIG074102: hypothetical protein
240	46	816	+	FIG00638373: hypothetical protein
240	893	2005	+	Mobile element protein
291	134	2362	-	hypothetical protein
291	2359	3540	-	FIG00715976: hypothetical protein
291	3677	4537	+	COGs COG2378
291	4914	5063	+	hypothetical protein
291	5102	7306	-	site-specific recombinase, phage integrase family
291	7296	8303	-	possible integrase
362	269	1273	-	Mobile element protein
364	289	570	+	hypothetical protein
590	55	372	-	Type I restriction-modification system, restriction subunit R (EC 3.1.21.3);Ontology_term=KEGG_ENZYME:3.1.21.3

Supplementary Table 4.12| List of Annotated Genes identified in the deletion of lineage A.

Category	Annotated genes
Fatty Acid Synthesis	(3R)-hydroxymyristoyl-[ACP] dehydratase (EC 4.2.1.-)
	3-hydroxydecanoyl-[ACP] dehydratase (EC 4.2.1.60)
	3-oxoacyl-[ACP] reductase (EC 1.1.1.100)
	3-oxoacyl-[ACP] synthase
	3-oxoacyl-[ACP] synthase (EC 2.3.1.41) FabV like
	Acyl carrier protein
	Acyl carrier protein (ACP1)
	FIG002571: 4-hydroxybenzoyl-CoA thioesterase domain protein
	FIG018329: 1-acyl-sn-glycerol-3-phosphate acyltransferase
	FIG143263: Glycosyl transferase / Lysophospholipid acyltransferase
	FIGfam138462: Acyl-CoA synthetase, AMP-(fatty) acid ligase
Carbohydrate and	D-galactarate dehydratase (EC 4.2.1.42)
Amino Acid Metabolism	D-galactarate dehydratase (EC 4.2.1.42)
	Tagatose-6-phosphate kinase AgaZ (EC 2.7.1.144)
	Putative O-methyltransferase
	Aspartate aminotransferase (EC 2.6.1.1)
	2-keto-3-deoxy-D-arabino-heptulosonate-7-phosphate synthase I alpha (EC 2.5.1.54)
Iron Scavenging	Glycosyltransferase IroB
	ABC transporter protein IroC
	Trilactone hydrolase IroD
	Periplasmic esterase IroE
	Outer Membrane Siderophore Receptor IroN
	Hemolysin activator protein precursor
	Putative large exoprotein involved in heme utilization or adhesion of ShlA/HecA/FhaA family
Bacteriocins and	antigen 43 precursor
Virulence	Colicin-E1*
	mannose-specific adhesin FimH
	Probable microcin H47 secretion/processing ATP-binding protein mchF (EC 3.4.22.-)
	MchC protein Putative F1C and S fimbrial switch Regulatory protein
	type 1 fimbrae adaptor subunit FimF
	type 1 fimbriae anchoring protein FimD
	type 1 fimbriae major subunit FimA
	type 1 fimbriae protein FimI2C unknown function
	YeeV toxin protein

Category	Annotated genes
Other	Transposase (X4)
	Putative Transposase
	Mobile element protein
	Integrase IS, phage, Tn: Transposon-related functions
	Putative metal chaperone, involved in Zn homeostasis, GTPase of COG0523 family
	entry exclusion protein 2
	membrane: Transport of small molecules: Cations
	FIG021862: membrane protein, exporter
	FIG027190: Putative transmembrane protein
	putative membrane protein (X2)
	putative secretion permease
	putative regulatory protein
	NgrB
	RNA-Arg-TCT

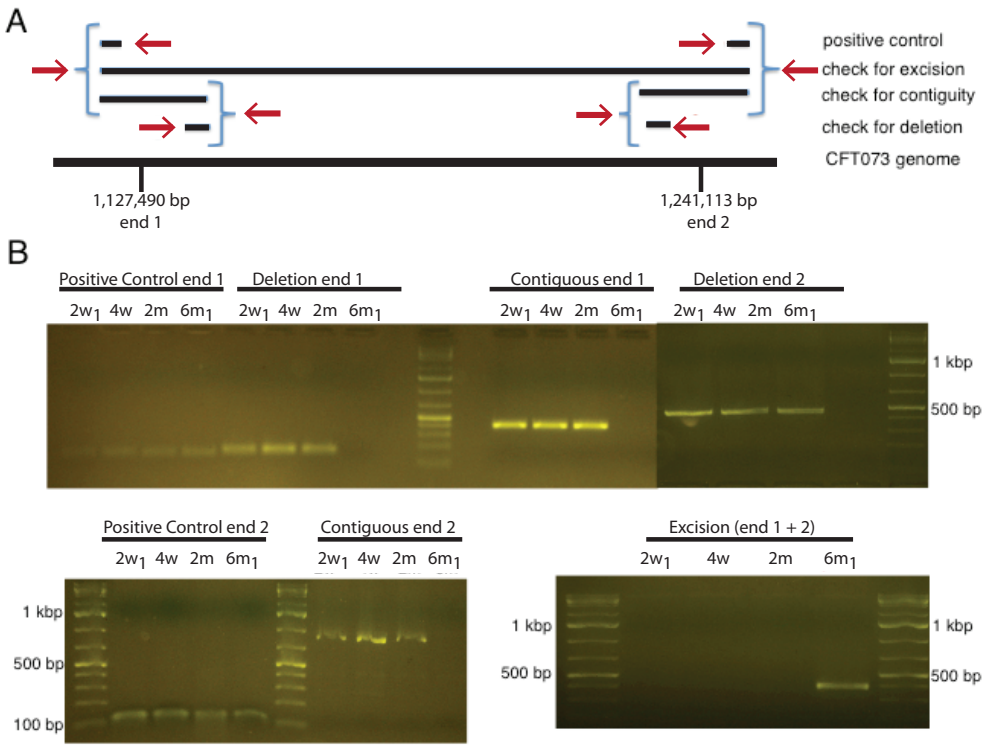
Supplementary Table 4.13| List of Annotated Genes identified in the deletion of lineage B.

Category	Annotated genes
Virulence	putative regulator PapX protein
	PapG protein
	PapF protein
	PapE protein
	Fimbrial adapter papK precursor
	Periplasmic fimbrial chaperone
	Fimbriae usher protein StfC
	minor pilin subunit PapH
	major pilin subunit PapA
	PapI protein
Iron Scavenging	TonB-dependent receptor; Outer membrane receptor for ferrienterochelin and colicins
Other	Mobile element protein (X9)
	conserved hypothetical protein; putative exported protein
	hypothetical protein (X4)
	FIG00638040: hypothetical protein
	FIG00638658: hypothetical protein
	FIG00640659: hypothetical protein
	FIG01070050: hypothetical protein
	FIG00641173: hypothetical protein

Supplementary Table 4.14| Annotations of the pNK117-2 plasmid. Annotations were done via RAST, but modified to include homologous genes from pOLA52 and from the *Raoultella planticola* RJA274 plasmid.

Start	Stop	Strand	Annotation
830	1093	-	IncN plasmid KikA protein
1129	1368	-	hypothetical protein
1539	3374	-	Type IV secretion system protein VirD4
3377	4408	-	ATPase provides energy for both assembly of type IV secretion complex and secretion of T-DNA complex (VirB11)
4410	5615	-	Inner membrane protein of type IV secretion of T-DNA complex2C TonB-like2C VirB10
5612	6541	-	Forms the bulk of type IV secretion complex that spans outer membrane and periplasm (VirB9)
6546	7259	-	Inner membrane protein forms channel for type IV secretion of T-DNA complex (VirB8)
7249	7377	-	hypothetical protein
7505	8635	-	Integral inner membrane protein of type IV secretion complex (VirB6)
8644	8931	-	IncQ plasmid conjugative transfer protein TraG
8932	9690	-	Minor pilin of type IV secretion complex (VirB5)
9701	12331	-	ATPase provides energy for both assembly of type IV secretion complex and secretion of T-DNA complex (VirB4)
12476	12772	-	Major pilus subunit of type IV secretion complex2C VirB2
12749	13397	-	virB1, homologous to pOLA52 pilx1
13572	13934	-	hypothetical protein
13571	14090	-	actX, homologous to pOLA52
14443	15609	-	IncQ plasmid conjugative transfer DNA nicking endonuclease TraR (pTi VirD2 homolog)
15612	16226	-	DNA distortion protein 1
15611	16157	-	taxA
16273	16422	-	hypothetical protein
16483	16755	-	FIG01048616: hypothetical protein
16987	17166	-	hypothetical protein
17209	17538	-	YagA protein
17631	17846	-	FIG01048886: hypothetical protein
17836	18081	-	FIG01047979: hypothetical protein
18126	18449	-	FIG01047054: hypothetical protein
18595	18876	-	RelE/StbE replicon stabilization toxin
18866	19117	-	RelB/StbD replicon stabilization protein (antitoxin to RelE/StbE)
19209	19340	+	hypothetical protein
19901	20032	-	hypothetical protein
20349	21185	+	PI protein (pir)
21225	21668	+	bis, homologous to RJA274 NDM-1 producing plasmid
21856	22212	+	DNA distortion protein 3 (ddp3)

Start	Stop	Strand	Annotation
22209	23186	-	hypothetical protein
23211	23492	-	hypothetical protein
23592	24254	-	Chromosome partitioning protein ParA (parA)
24635	25603	+	Resolvase (res)
25758	25874	+	hypothetical protein
25942	26082	-	hypothetical protein
26268	26402	-	hypothetical protein
26473	26658	-	hypothetical protein
26616	26894	-	hypothetical protein
26912	27166	-	hypothetical protein
27277	28101	-	zinc metalloproteinase Mpr protein (mpr)
28104	28337	-	hypothetical protein
28393	28569	-	hypothetical protein
28547	28954	-	hypothetical protein
29010	29474	-	DNA-binding protein H-NS (hns)
29490	29696	-	Haemolysin expression modulating protein
29693	31846	-	DNA topoisomerase III (EC 5.99.1.2);Ontology_term=KEGG_ENZYME:5.99.1.2 (topB)
31851	32426	-	FIG01045518: hypothetical protein
32430	33692	-	Cell division protein FtsH (EC 3.4.24.-);Ontology_term=KEGG_ENZYME:3.4.24.- (ftsH)
33774	34265	-	DNA topoisomerase III (EC 5.99.1.2);Ontology_term=KEGG_ENZYME:5.99.1.2
34694	34951	-	hypothetical protein
35082	35309	-	hypothetical protein



Supplementary Figure 4.1 | PCR reactions to confirm the deletion in lineage A isolates from 6 months and show it was an excision. A. Overview of the PCR primer strategy. Contigs were aligned to CFT073 to identify the order of the contigs involved in the deletion. PCR primers were designed to confirm that the contigs were contiguous prior to the deletion (samples $2w_1$, 4w, and $6m_1$). PCR primers were also designed to distinguish isolates without the deletion (samples $2w_1$, 4w, and $6m_1$), and that the deletion identified in lineage A from 6 months was an excision (demonstrated here with isolate $6m_1$). B. PCR products separated via gel electrophoresis to view PCR results. Contig alignment to CFT073 did order the contigs correctly, as shown in the 'Contiguous end 1' and 'Contiguous end 2' labelled gels, and that the isolates $2w_1$, 4w, and $6m_1$ did not have the deletion. Furthermore, the deletion was an excision that occurred in the lineage A isolates from 6 months, demonstrated here using $6m_1$.

CHAPTER 5

Pathogenic and commensal *Escherichia coli* lineage dynamics *in situ* of the gut during antibiotic treatment

Heidi K. Gumpert[1], Nahid Karami[2], Agnes E. Wold[2], Ingegerd Adlerberth[2], Morten O. A. Sommer[1,3]

1) Department of Systems Biology, Technical University of Denmark, DK-2800 Lyngby, Denmark.
2) Department of Infectious diseases, Section Clinical bacteriology, Sahlgrenska academy, S-405 30 Göteborg, Sweden.
3) Novo Nordisk Foundation Center for Biosustainability, DK-2900 Hørsholm, Denmark.

5.1 Abstract

The human gut is considered a hub for horizontal gene exchange, and has been shown to be a reservoir of antibiotic resistance genes. The potential for bacteria in the gut, many of which are opportunistic pathogens, to acquire resistance genes or virulence determinants and lead to infections that are more difficult to treat is of high concern. One of the commensal microbes of the human gut microbiota is *Escherichia coli*, and is also a well known pathogen. Here, we follow co-existing *E. coli* lineages *in situ* of the gut through an infants first year of life. We identify that it was an *E. coli* lineage colonizing the gut that caused a urinary tract infection in the infant and transferred a conjugative plasmid harbouring a beta-lactamase resistance gene to another co-existing *E. coli* lineage *in situ* of the gut. This uropathogenic *E. coli* strain was found to additionally harbor a disseminated plasmid that is implicated in virulence. While the transconjugant strain was not sampled again after receiving the resistance plasmid, the uropathogenic strain colonized the gut for the entire year of the study. Through genome sequencing of these co-existing *E. coli* we observed genomic events in these lineages, and catch a glimpse of the dynamics of *E. coli* lineages in the gut.

5.2 Introduction

The human gut is home to one of the densest microbial ecosystems, the human gut microbiota, and this microbial community plays an important role in human health and physiology. Among the commensal residents of the gut microbiota, there are opportunistic pathogens that can cause infection following perturbation to their host [1]. Additionally, the gut can serve as a reservoir of bacteria able to cause infection at other body sites. One such example is *Escherichia coli*, which is a commensal microbe of the human gut microbiota, but is also a well known pathogen with strains causing both intestinal and extra-intestinal infections. It is well established that the majority of urinary tract infections (UTI) are caused by *E. coli* strains inhabiting the gut, although other predisposing factors and transmission routes have been recorded [2, 3].

The human gut is considered a hub for horizontal gene exchange. It has been shown to be a reservoir of antibiotic resistance genes [4], and could similarly be a reservoir of virulence genes [5]. It is not only bacteria colonizing the gut that may partake in horizontal gene transfer, but also bacteria that are only transiting through

the gut. Acquisition of antibiotic resistance genes could yield a pathogenic strain more difficult to treat, while acquisition of virulence determinants could render a strain pathogenic, or more pathogenic. Transiting bacteria involved in horizontal gene transfer could leave the gut with new antibiotic resistance or virulence determinants and contaminate the sites they normally colonize and/or infect with the newly acquired resistance or virulence genes [5].

Investigations into the dynamic nature of the gut microbiota at the level of strain resolution have included both observations of horizontal gene transfer, as well as strain abundance and turn over. A few phenotypic reports exist documenting horizontal gene transfer of antibiotic resistance genes between the constituents of the human gut microbiota [6, 7]. In these reports, transfer of resistance genes was observed in connection with antibiotic treatment and detected via a change in the antibiotic resistance susceptibility of the strain. Non-cultivation based methods have been developed to investigate the abundances of strains during colonization of the gut [8, 9, 10]. In these studies, time-series metagenomic sampling is processed in a manner that yields strain resolution, with near-complete and partially-complete genomes assembled for microbial constituents comprising as little as 1% and 0.05%, respectively [9]. However, as this method is based on employing strain abundance profiles to group sequences belonging to the same strain, it is not particularly well suited to observe horizontal gene transfer events.

In this study, we examine co-existing *E. coli* lineages sampled from an infant over the first year of life. This set of isolates was initially selected to genomically investigate a reported transfer of a beta-lactamase antibiotic resistance gene between co-existing *E. coli* lineages *in situ* of the gut [7]. As this infant developed a UTI within the first month of life, we additionally wanted to link the uropathogenic strain back to the lineage dynamics of the source strain in the gut. By full genome sequencing this set of isolates, we gain insights into the dynamic nature of *E. coli* strains co-existing in the gut.

5.3 Results

Study Material

The current study material was obtained from an infant enrolled in the ALLERGYFLORA study. The aim of the ALLERGYFLORA study was to investigate the relationship between intestinal colonization patterns to the development of allergies

later on in life [11]. Samples from the infants enrolled in this study to investigate the colonization of the gut microbiota were collected at 2 days after birth via rectal swab, and 1, 2, and 4 weeks and 2, 6, and 12 months after birth via stool sample. *Escherichia coli* was quantitatively cultured on Drigalski agar plates, with various morphotypes enumerated separately. Lineage identities of the *E. coli* colonies were determined via random amplified polymorphic DNA (RAPD) typing.

E. coli Lineage Sampling

In total, two *E. coli* lineages were sampled from fecal matter sampled from the infant over the first year of life (Fig. 5.1) [7]. Sampling time points were at 2, 9, 16 and 32 days, and 2, 6 and 12 months of age. *E. coli* lineage A was recovered at all the sampling time points. A second *E. coli* lineage, termed lineage B, was only first sampled at the 9 day sampling point. Lineage B was sampled again at 16 days, and lastly at 32 days.

Isolates were screened for their antibiotic susceptibility profiles using the disc diffusion method (Oxoid, Sollentuna, Sweden). Antibiotics to which an isolate was not susceptible to were followed up by minimum inhibitory concentration (MIC) testing [7]. The lineage A isolates were found to be highly resistant to ampicillin (MIC $\geq$ 256mg/L) (Fig. 5.1). The first two lineage B isolates at 9 and 16 days were initially susceptible to ampicillin. However, the lineage B isolate at 32 days had a change in its antibiotic susceptibility profile and was discovered to also be highly resistant to ampicillin (MIC $\geq$ 256 mg/L). Lineage A isolates were also resistant to piperacillin, but isolates collected from 32 days to 12 months were resistant to a lesser degree [7].

An additional isolate from a urinary tract infection is also included in the study. At 8 days of age, the infant boy was admitted to hospital due to a suspected urinary tract infection. He was administered trimethoprim for 5 days. A urinary culture taken at 11 days of age yielded *E. coli* at 100 000 colony forming units (CFU) per milliliter. However, due to enterococci in addition to *E. coli* in a subsequent urinary sample, the antibiotic treatment was changed to intravenous ampicillin for 5 days, followed by amoxicillin perorally for an additional 8 days. Lastly, the infant was administered trimethoprim prophylactically for the next 7 months.

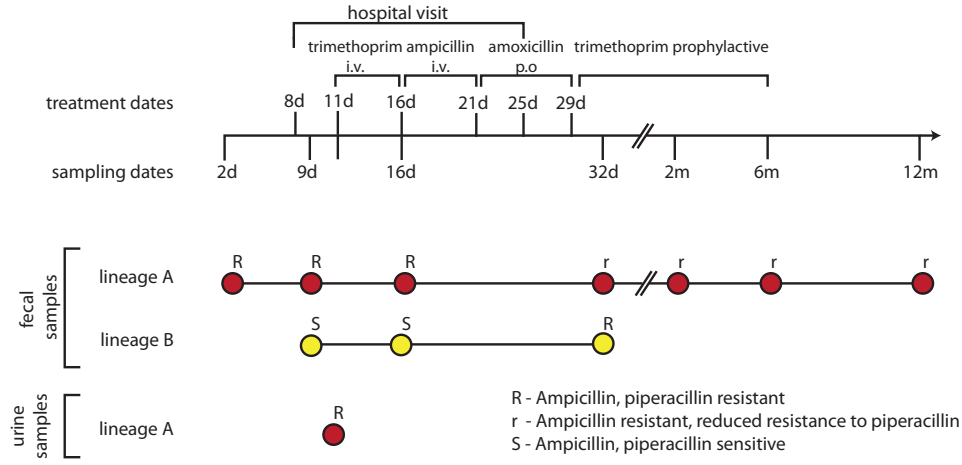


Figure 5.1 | Sampling of the co-existing *E. coli* lineages from the gut microbiota, and the UTI isolate, and antibiotic treatment of the infant.

Genome Sequencing of the Isolates

E. coli isolates obtained from fecal samples and from the urinary sample were genome sequenced. One isolate from each lineage for each sampling time was selected, except for A lineage at 2 and 6 months where two isolates were selected for sequencing. At least one isolate from each lineage was sequenced using an Illumina MiSeq yielding 150 base-pair (bp) paired-end reads. Specifically, lineage A isolate from 2 days and 2 and 6 months, labelled 2d, $2m_2$ and $6m_1$, respectively, lineage B isolate from 32 days, and the UTI isolate were selected. Representative genomes for the lineage or isolation site were assembled from the reads from the A lineage at 2 days, the B lineage at 32 days, and the UTI isolate. All other isolates were sequenced on an Illumina HiSeq yielding 70bp single-end reads. Comparisons between isolates were conducted to ensure that the isolates selected as lineage and site representatives were indeed representative of the lineage and isolation site (see Materials and Methods).

The total genome length, including plasmids, for the lineage A is 4,907,304 bp (4,750,822 bp without plasmids) from a total of 90 contigs. The N50 of the lineage A assembly is 141,096 bp. Two conjugative plasmids were identified in all lineage A isolates. One plasmid was identified as an IncX plasmid 42 kb in length, and the other as a 114 kb plasmid identified with FIB, FII(29), Col(156) replicons, typed as

F29:A-:B10 in the IncF pMLST typing scheme [12, 13] (Supplementary Table 5.3).

The total genome length, including plasmids, for lineage B is 5,450,674 bp, from a total of 239 contigs. The N50 of the lineage B assembly is 123,856 bp. One plasmid of approximately 88 kb was found in all three lineage B isolates. *In silico* replicon typing revealed that this plasmid belonged to the p0111 replicon group, that was named as such as no RepA protein from a previously assigned Inc group was homologous to the *repA* gene from plasmid p0111 [12]. Two other replicon groups identified via the PlasmidFinder tool were B/O/K/Z and Col(MG828). With the assembly it was unfortunately not possible to characterize these plasmids, but the coverage depth data suggests that the replicons are found on two distinct plasmids, with the B/O/K/Z plasmid on the order of 110 kb, and the Col(MG828) plasmid on the order of 15 kb.

Genomic Confirmation of Lineage Identities

To confirm the lineage identities of the collected isolates, we first confirmed that the two lineages identified via RAPD typing were in fact distinct lineages. Using the single nucleotide polymorphism (SNP) calling pipeline (see Materials and Methods), there were on the order of 98,000 SNPs between the two representative isolates for each lineage. Furthermore, when assessing the shared total genetic content, only 77% and 85% was shared in common between the A and B lineage.

To assess the relationship of the lineage isolates, SNP-based phylogenetic trees were constructed (Fig. 5.2, Supplementary Table 5.1). For lineage A, no SNPs were detected between the isolates collected at days 2, 9 and 16, and only one SNP detected between them and the isolate collected at 32 days. Two isolates were included that were sampled at 2 months, and these shared 5 SNPs compared to the 32 day isolate, but then branched with 13 and 14 SNPs from this point, for the $2m_1$ and $2m_2$ isolates respectively. Two isolates were also included that were sampled at 6 months. The $6m_1$ isolate is located off the branch from the $2m_2$ isolate. The $6m_2$ is distantly related to both 2m isolates, as seen from the branching off point only 3 SNPs after the 32 day isolate. The isolate collected at 12 months is most closely related to the $6m_2$ with a total of 158 SNPs between them, or 136 SNPs from the 12 month isolate to their common ancestor.

There were no SNPs detected in the three isolates from lineage B.

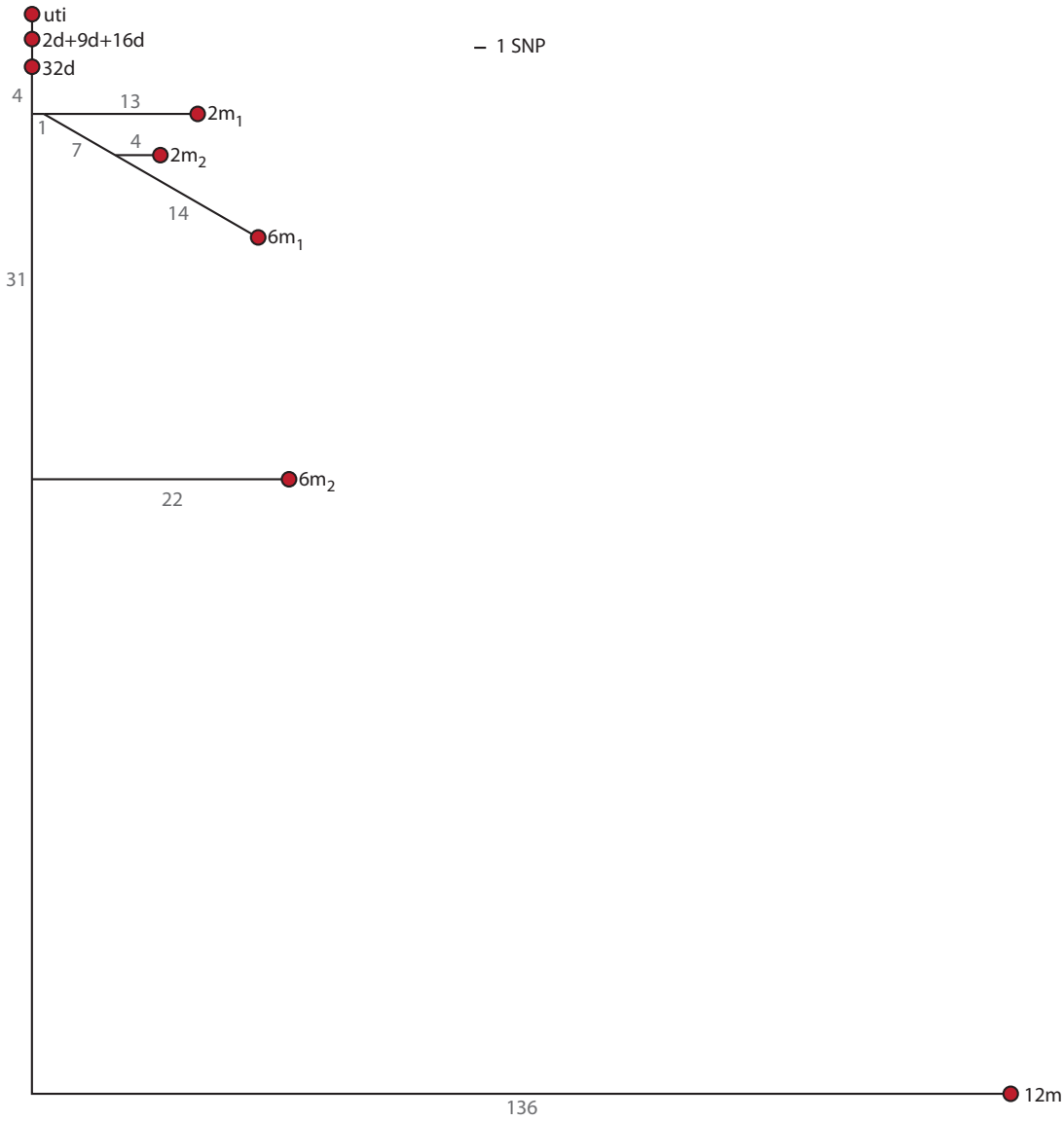


Figure 5.2 | SNP tree of the *E. coli* lineage A from the gut microbiota with the UTI isolate. The numbers beside branches indicate the length of the branch.

E. coli colonizing the intestine caused the UTI

The *E. coli* UTI isolate was compared to *E. coli* isolates from the infant's stool samples. Only one SNP was found when compared to the A lineage isolates collected at 2, 9 and 16 days (Fig. 5.2). Given the very high similarity, the A lineage colonizing the gut microbiota was the likely source of the UTI.

When analyzing the plasmid content of the A lineage, contigs from one of the plasmids were found to have very high sequence identity to a few plasmids deposited in NCBI's Genbank. The SNP calling pipeline was used to compare similarity to this plasmid, termed pNK29-2 (Table 5.1). In particular, the pUTI89 plasmid was found to have only 7 SNPs difference to the pNK29-2 plasmid, and aligning the contigs from this plasmid showed that there were no additional insertions.

Interestingly, these three highly similar plasmids were all isolated from various pathogenic *E. coli* strains. Strain UTI89, harbouring the pUTI89 plasmid, is an uropathogenic *E. coli* (UPEC) strain isolated from a patient with an acute bladder infection [14, 15]. Strain UM146, harbouring the pUM146 plasmid, is an adhesive invasive *E. coli* (AIEC) strain isolated from the ileum of a Crohn's disease patient [16, 17]. Strain EC14, harbouring the pEC14_144 plasmid, is also an UPEC strain isolated from a UTI [18].

Genome sequences are available for UTI89 and UM146, but when compared to lineage A, they did not have very high similarity. Only 88.4% of the 5,065,740 bp UTI89 genome and 89.5% of the 4,993,012 bp UM146 genome was covered when aligning reads from lineage A. Additionally, the serotypes of UTI89, EC14 and lineage A all varied, suggesting that different host *E. coli* strains harboured a very similar plasmid. Furthermore, the geographical locations where these strains were isolated varied, from the US, Canada and Sweden, suggesting that these pUTI89-like plasmids are globally disseminated (Table 5.2).

Wijetunge *et al.* have recently reported sequencing a plasmid with 99% identity to pUTI89, isolated from a neonatal meningitis *E. coli* (NMEC) strain, RS218 [19]. Interestingly, this plasmid shared only 41 - 46% similarity to other previously sequenced plasmids from NMEC strains [19]. The sequence of this plasmid, pRS218, was unfortunately not publicly available yet at the time of writing, but would also be expected to have high similarity to pNK29-3. However, RS218 has the same serotype as UTI89, O18:K1:H7, so it will be interesting to see the similarity and differences of these two pathogenic strains when the group publishes the genome of RS218.

Table 5.1 | The number of SNPs and INDELs identified when compared to the sequences of the plasmid listed below.

Plasmid Name	No. SNPs	No. INDELs	Plasmid Length (bp)
pUTI89	7	0	114,230
pUM146	34	11	114,550
pEC14_144	38	7	114,222

Table 5.2 | Plasmid host name and serotype, type of pathogenic *E. coli* (UPEC: uropathogenic *E. coli*, AIEC: adherent, invasive *E. coli*, NMEC: neonatal, meningitis *E. coli*) and geographical location of isolation.

Plasmid Name	Host Name	Serotype	Pathogen Type	Geographical location
pNK29-2	lineage A	O134: K1:H31	UPEC	Sweden
pUTI89	UTI89	O18: K1:H7	UPEC	USA
pUM146	UM146	not determined	AIEC	Canada
pEC14_144	EC14	O6:H31	UPEC	USA
pRS218	RS218	O18: K1:H7	NMEC	USA

When Chen *et al.* sequenced UTI89 and the pUTI89 plasmid, they identified several characteristics in common with the F plasmid [20]. They described a full *tra* operon for conjugative transfer and plasmid replication and regulation genes such as *repB*, *repA6*, *repA1*, *ccdA*, and *ccdB* to suggest it is replicated in low copy numbers, which is in agreement with our coverage depth of the plasmid relative to the genome [20]. Additionally, several genes on the plasmid are related to pathogenesis, including the *cjrABC* operon, which is believed to encode proteins involved in iron uptake, and *cjrB* strains sensitive to colicin J, and the *senB* gene, encoding an enterotoxin [20, 21, 22]. These four genes are conserved between the virulence plasmid from the enteroinvasive *E. coli* (EIEC) O164 strain [20, 21]. Six open reading frames (ORFs) showed high identity to the UTI89 chromosome, in addition to other sequences encoding proteins of non-plasmid origin. These results suggest that F-like plasmids, such as pUTI89, undergo frequent recombination [20].

After Debroy *et al.* sequenced the pEC14_144 plasmid, apart from noting the striking sequence similarity to the pUTI89 plasmid, they also found it very similar

to the p1ESCUM plasmid [18]. When comparing the pEC14_144 plasmid to the p1ESCUM plasmid from the UMN026 UPEC strain, they found that there were only 46 SNPs between p1ESCUM and pEC14_144, but they identified three insertions. The largest of these insertions harboured the *tetRACD* tetracycline resistance genes.

Out of the 7 SNPs that were identified in pNK29-2 relative to pUTI89, only two of these were non-synonymous changes (Supplementary Table 5.2). One of these was in the *rsvB* gene, a resolvase, and in the *traE* gene, a conjugal transfer protein for F pilus assembly.

Cusumano *et al.* have studied the role of pUTI89 in urinary tract infections caused by UTI89 [21]. In an *in vitro* model, they did not observe any differences between UTI89 and a plasmid-cured UTI89 strain when comparing growth, type 1 pilus expression, or biofilm formation. However, in an *in vivo* mouse model for UTI, they did observe significant differences in the early aspects of infection for the pUTI89-cured strain. Specifically, they observed a decrease in bacterial invasion, a decrease in the number of bacteria in the bladder, and a decrease in the formation of biofilm-like intracellular bacterial communities. Similarly, Wijetunge *et al.* compared RS218 and plasmid-cured RS218 strains, but found an attenuation of the plasmid-cured strain for both *in vitro* invasion potential into human cerebral microvascular endothelial cells and *in vivo* with infected rat pups with regards to mortalities, histopathological lesions in the brain tissue, and bacterial recovery [19]. Apart from knowing that lineage A had successfully caused a UTI, the exact role of the pNK29-2 plasmid in causing the infection can only be speculated given the different *E. coli* host.

Genomic Confirmation of Antibiotic Resistance Genes Transfer *in situ* of the gut

Karami *et al.* have previously reported that an observed change in the antibiotic susceptibility profile of lineage B [7]. When lineage B was sampled at 9 and 16 days, it was susceptible to ampicillin. At 32 days, lineage B was found to be resistant to ampicillin (Fig 5.1). In addition to phenotypic evidence of the antibiotic susceptibility change, evidence for the transfer included DNA sequencing yielding the same bla_{TEM-1b} variant. Additionally, a plasmid restriction fragment length polymorphism showed similar DNA fragments, from a plasmid approximately 40 kb in size, termed pNK29.

Here we have determined the sequence of the pNK29 conjugative plasmid of 42.2

kb. *In silico* analysis revealed that the plasmid belongs to the IncX1 incompatibility group [12]. Annotation of the plasmid was initially done by using RAST [23], which predicted that the plasmid encodes a total of 56 proteins. We compared pNK29 to pOLA52, a well-annotated IncX1 plasmid [24], and to pRPEC180_47 [25], identified via BLAST as the most closely related plasmid based on coverage and identity at 83% and 96%, respectively (Fig. 5.3). Using pOLA52 to identify homologues in pNK29, we refined the RAST annotation to include an additional plasmid replication initiator protein, *pir*, and a replication initiation auxiliary protein, *bis*. Additionally, *actX* and *pilx7* were identified based on homology. Similar to other plasmids of the IncX compatibility group, this plasmid also followed the generalized structure of *pir-bis-par-hns-topB-pilX-actX-taxCA* (Fig. 5.3) [25]. In the genetic load area of IncX1 plasmids, between the *par-res* genes and the *hns-topB* genes [25], this plasmid contains a total of 11 predicted ORFs. The bla_{TEM-1b} gene is located in this region. The full list of annotations can be found in Supplementary table 5.4.

Karami *et al.* had documented a change in the resistance phenotype of lineage A isolates collected from 32 days and onwards. The isolates had a reduced resistance to piperacillin due to a single nucleotide difference in the promoter region of the bla_{TEM-1b} gene on pNK29, converting the weak *P3* promoter into a strong *Pa/Pb* [7]. This SNP is present on the phylogenetic tree as the single nucleotide difference between the 2d, 9d and 16d isolates and the 32d isolate (Fig 5.2).

Lineage A Dynamics

The analysis of the isolates collected from lineage A revealed two differences in the genomic content between the isolates. The earlier collected isolates appear to have an integrated phage, whereas the later collected isolates had a small, cryptic plasmid.

When comparing the isolates to the representative genome of the A lineage, the $6m_2$ and 12m isolates appeared to be missing approximately 54 kb. Aligning the contigs confirmed that there was extra genomic content in the isolate from 2 days, compared to the $6m_2$ and 12m isolates. Inspection of the annotations in this region led to the suspicion that the 54 kb region was an integrated phage. Annotated genes in the regions included multiple phage tail fiber, assembly and tail length tape-measure proteins, DNA transfer proteins, phage terminase, endopeptidase, lysozyme, and integrase proteins, and an additional 62 hypothetical proteins. Aligning the flanking sequences of this prophage-associated region and the contigs from the $6m_2$ and 12m

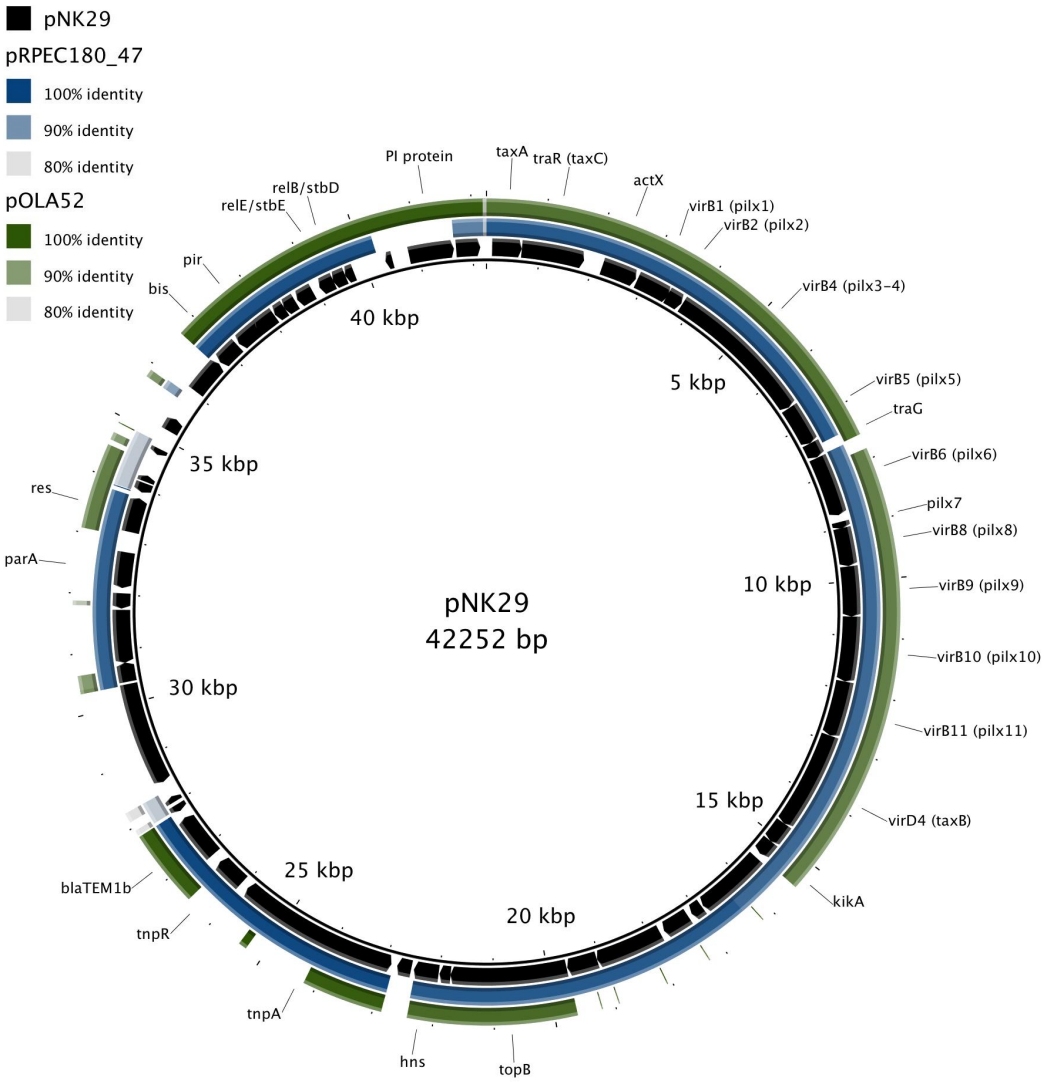


Figure 5.3 | Plasmid pNK29 compared to other IncX1 plasmids pRPEC180_47 (middle ring, blue) and pOLA52 (outer ring, green). Open reading frames (ORFs) identified on pNK29 are drawn in the inner most ring in black, with arrows indicating the reading direction. Annotations for some ORFs are labeled outside of the rings, with the gene name in pOLA52 in brackets.

isolates revealed a potential 12 bp *attR* and *attL* sequence, TCCAGACAGGAC.

To determine if this prophage-associated region could be an intact phage, we used the PHAST web service [26]. However, PHAST predicted that only the latter 33.4 kb of this 54.2 kb region as an integrated phage, but it did predict this region to contain an intact, functional prophage. A BLAST search of this suspected phage region yielded only two hits with greater than 2% coverage. One of the hits was a recently deposited 288 kb scaffold from the FHI99 *E. coli* strain, yielding 90% coverage with 94% identity. A second hit was from the complete assembled UMNK88 *E. coli* genome, yielding 80% coverage with 89% identity. Shepard *et al.* had in the analysis of the UMNK88 genome identified this homologous 55 kb region as prophage-associated [27]. PHAST analysis of both the scaffold from FH199 and the UMNK88 genome identified the entire homologous region as an intact prophage [26]. All three prophage regions were determined to be most similar to the 49 kb *Salmonella* phage vB_SosS_Oslo, a member of the Siphoviridae family with long non-contractile tails.

In the $6m_2$ and 12m isolates, a small cryptic plasmid was found, termed pNK29-3. This 2545 bp plasmid has a low GC content of 33.4% and contains two ORFs both predicted as hypothetical proteins by RAST annotation. By comparing the coverage depth of the plasmid to the average coverage depth of the genome, we estimate the copy number to be around 9 plasmids per cell. *In silico* typing of pNK29-3 revealed that it belongs to the ColE1-like group of small plasmids (Supplementary Table 5.3). The probe for this classification was based on the mobility protein sequence from the pIGMS31 plasmid [12]. The pIGMS31 plasmid was isolated from a pathogenic *Klebsiella pneumoniae* strain and is a narrow host range plasmid that can be maintained in Gammaproteobacteria [28]. Comparing plasmid pNK29-3 to pIGMS31 yields 63% coverage by pIGMS31 with 95% identity. The region of homology includes an identified *oriT* site, and a mobilization protein, which Smorawinska *et al.* have demonstrated is functional in a wider range of hosts than the replication system (Fig. 5.4) [28]. The other ORF on pNK29-3 did not share any similarity to the predicted *rep* gene in pIGMS31. One highly similar plasmid to pNK29-3 was identified by BLAST in NCBI genbank was the pEA1 plasmid isolated from a *Pantoa agglomerans* strain (formerly *Enterobacter agglomerans*). Plasmid pEA1 is the same length as pNK29-3, covers 100% of pNK29-3 and shares 96% identity. However, this plasmid is neither annotated nor characterized, so it is not possible to predict how pNK29-3 initiates replication based on sequence homology.

These two genetic content differences can be observed in the SNP-based tree (Fig.

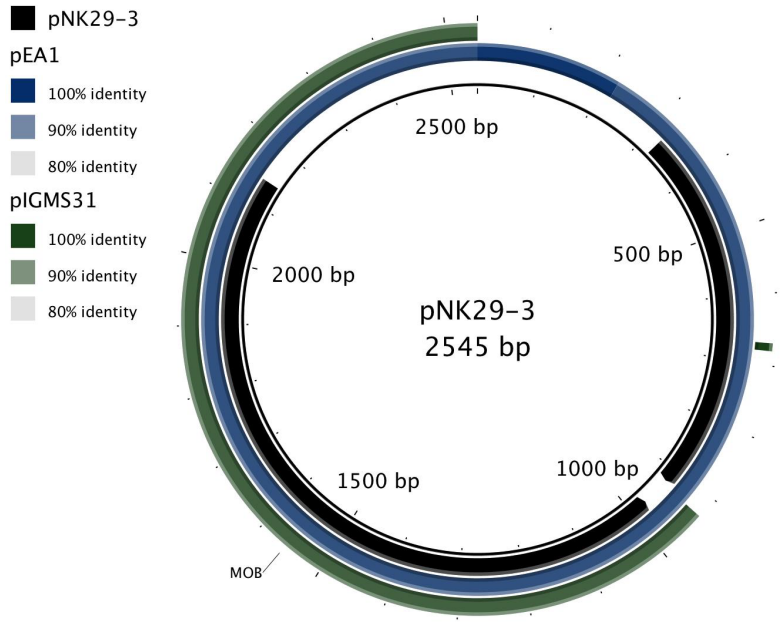


Figure 5.4 | Plasmid pNK29-3 compared to other ColE1-like plasmids pEA1 (middle ring, blue) and pIMGS31 (outer ring, green). Open reading frames (ORFs) identified on pNK29-3 are drawn in the inner most ring in black, with arrows indicating the reading direction.

5.2). The isolates with the integrated phage, 2d, 9d, 16d, 32d, $2m_1$, $2m_2$, 6_1 and the UTI isolate, are located together with a maximum of 28 SNPs detected between them. A long branch after the 32d isolate separates the isolates with the integrated phage, with the $6m_2$ and 12m isolates without the phage, but with the small, cryptic plasmid, pNK29-3. From this branch after the 32d isolate, there are 53 SNPs to the nearest isolate without the phage, the $6m_2$ isolate, and 177 SNPs to the 12m isolate. Based on the phylogenetic tree, it is likely that either the prophage was cleared or plasmid was acquired at some point between 32 days and 2 months, even though an isolate without the phage or with pNK29-3 was not sampled at 2 months.

5.4 Discussion

By genome sequencing this collection of isolates from an infant over the first year of life, we are able to observe complex dynamics of these co-existing *E. coli* lineages *in situ* of the gut.

We have been able to confirm that it was *E. coli* lineage A that colonized the gut causing a UTI in this infant. We identified a 114 kb plasmid that was strikingly similar to other previously sequenced plasmids from pathogenic *E. coli* strains. This plasmid in a host with a different genetic background has been shown in a mouse model to play a role in the initial stages of infection [21], however we can only speculate as to the role of this plasmid in the lineage A ability to cause a UTI infection. Nevertheless, observing such a similar virulence-involved plasmid again in a different genetic background and in different geographical regions does raise questions how this plasmid was able to disseminate and be maintained successfully in different hosts. Assessing the role of this plasmid on the ability of its host to colonize the gut may lead to evidence explaining why it has been maintained in the various hosts it has been found in. In this study, lineage A remained colonized in the gut the entire duration of the study, from 2 days to one year after birth.

We have also been able to confirm via genome sequencing that it was also lineage A that transferred the pNK29 plasmid to the co-existing lineage B *in situ* in the gut. This plasmid conferred resistance to ampicillin and reduced susceptibility to amoxicillin/clavulanic acid, which would have been beneficial for survival in the gut while the infant was undergoing treatment [7]. This benefit was also observed in a dramatic increase in the population counts of lineage B in the microbiota before and after the transfer [7].

While the isolates collected for the ALLERGYFLORA were originally selected to serve as a lineage representative for each time point to answer questions regarding colonization patterns, we have been able to catch a glimpse into the multi-clonal environment of the human gut. By including only one extra isolate for two time points, 2 and 6 months, we are able to observe two different trajectories of one *E. coli* lineage. This intra-lineage difference is most pronounced with the two isolates collected at 6 months; $6m_1$ most related to the $2m_2$ isolate containing the prophage whereas the $6m_2$ isolate branched off much earlier after 32 days and does not contain the phage, but instead contains a small cryptic plasmid. While sequencing isolated single colonies has proven to be very powerful in identifying genomic events, such as

the loss of the integrated phage or acquisition of plasmids, sequencing more isolates or further refinement of current time series metagenomic sample processing methods will lead to a better understanding of the dynamics of the gut microbiota.

5.5 Materials and Methods

E. coli Lineage Sampling and Collecting of Isolates

Samples to investigate the colonization of the gut microbiota of the infants enrolled in this study were collected at 2 days after birth via rectal swab, and 1, 2, and 4 weeks and 2, 6, and 12 months after birth via stool sample. Stool samples were collected by the parents and brought to the lab and processed within 24 hours. Samples were quantitatively cultured for all major groups of facultative and anaerobic bacteria. *E. coli* was quantitatively cultured on Drigalski agar plates, with various morphotypes enumerated separately. E. coli isolates were speciated via API 20E identification strips (bioMériuex, France). Lineage identities of the *E. coli* isolates were initially determined via random amplified polymorphic DNA (RAPD) typing, and confirmed via pulsed-field gel electrophoresis (PFGE).

Genome Sequencing

Genomic DNA was isolated using an UltraClean®Microbial DNA Isolation Kit (Mo-Bio Laboratories, Inc., California). Sequencing libraries for Illumina HiSeq single-end sequencing were prepared by shearing DNA into 200bp fragments using a Covaris E210 and then constructing barcoded libraries. Illumina HiSeq sequencing was performed by Partners HealthCare Center for Personalized Genetic Medicine (Cambridge, Massachusetts). Illumina MiSeq paired-end sequencing libraries were prepared by using the Nextera XT (Illumina, California) for lineage A at 2 days, lineage B at 32 days, and the urinary tract isolate, and by using the TruSeq (Illumina, California) for lineage A at 2 months and lineage B at 9 days.

Sequence Analysis

Genomes for each sequenced isolate were assembled using Velvet (v1.2.09) [29] and annotated via RAST [23]. Various tests were conducted to ensure that the selected isolates were indeed representative for the lineage. First reads from the isolates were mapped onto the representative genome via Bowtie2 (2.1.0) [30], and single nucleotide

polymorphisms (SNPs) were enumerated via SAMTools (0.1.19) [31]. Additionally, to ensure that all isolates from within a lineage consisted of the same genomic content as the representative isolates, genomic areas lacking mapped read coverage were identified using BEDTools (2.17.0) [32]. This included both reads mapped onto the representative lineage, in addition to reads from the representative lineage mapped onto genomes assembled from reads of the other isolates. Scripts to assist in the analysis were written in python using Biopython libraries [33].

Coarse genome sequence comparisons were based on aligning reads from our study to other genome sequences, and assessing the number of SNPs via the SNP calling pipeline and/or assessing the coverage of the genome based on BEDTools [32]. This was done for comparison between lineages A and B, and for comparison of lineage A to UM146 (NC_017632.1) and UTI89 (NC_007946.1).

The pNK29 plasmid was assembled into a circular plasmid by aligning plasmid contigs together using MUMMER [34]. Contigs belonging to the plasmid were first identified in lineage B as the new genetic material of the isolate at 32 days, and then used to identify the corresponding contigs in lineage A genome. The introduction of possible errors in plasmid assembly were checked by aligning reads to the plasmid using Bowtie2 [30] and checking for SNPs and INDELs using SAMTools [31]. The RAST annotation for this plasmid was refined based on identifying homologous genes in pOLA52 (NC_010378.1) that were either missing or incorrect in pNK29. Additionally, the descriptions for the two mobile element proteins upstream of the bla_{TEM1b} gene were added on based on the homologous *tnpA* and *tnpR* genes in pRPEC180_47 (NC_019088.1).

Other plasmids were identified by first separating contigs with a copy number greater than one from the genomic contigs and then by grouping contigs with similar abundances together. The average coverage of each contig was determined using BEDTools [32]. Plasmid incompatibility grouping was done using the PlasmidFinder tool (http://cge.cbs.dtu.dk/services/PlasmidFinder) [12]. Homologous previously sequenced plasmids were identified using BLAST and the NCBI nt database [35]. Plasmid diagrams were created using the BLAST ring image generator (BRIG) [36], which uses BLAST to compare input sequences.

For pNK29-2, blastn searches of these plasmid contigs revealed 3 plasmids with very high identity: pUTI89 (NC_007941.1), pUM146 (NC_017630.1), pEC14_144 (NC_013175.1). These plasmids were downloaded and the SNP calling pipeline was used to assess SNP-based similarity. Additionally, the contigs with 2X-genomic cov-

erage were aligned to the pUTI89 plasmid to determine if the plasmid had any large insertions.

For pNK29-3, the small cryptic plasmid in the A lineage was identified based on the introduction of new genetic content in $6m_2$ and 12m isolates. Two similar plasmids identified via blastn were pIGMS31 (NC_011406.1) and pEA1 (DQ659147.1). Again, the RAST annotation was refined based on homolog to the pIGMS31 plasmid.

5.6 Acknowledgements

We would like to thank Mari Rodriguez de Evgrafov, Marius Faza and Andreas Porse for sequence library preparation, and Christian Munck and Lejla Imamovic for helpful discussions. This work was supported by the Danish Free Research Councils for Health and Disease, the European Union FP7-HEALTH- 2011-single-stage grant agreement 282004, EvoTAR, the Medical Faculty of the University of Göteborg (ALFGBG138401) and the Swedish Medical Research Council. M.O.A.S further acknowledges support from the Novo Nordisk Foundation, and the Lundbeck Foundation.

Author Contributions

NK, IA and AW designed the clinical study and isolated the *E. coli* strains. HG processed the sequencing data. HG and MS analyzed the sequencing data results and HG and wrote the manuscript with input from MS.

5.7 References

[1] Sam P Brown, Daniel M Cornforth, and Nicole Mideo. Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. *Trends in microbiology*, 20(7):336–342, July 2012.

[2] Evgeni V Sokurenko, Richard Gomulkiewicz, and Daniel E Dykhuizen. Source-sink dynamics of virulence evolution. *Nature Reviews Microbiology*, 4(7):548–555, July 2006.

[3] PD Brown and B Foxman. Pathogenesis of Urinary Tract Infection: the Role of Sexual Behavior and Sexual Transmission. *Current infectious disease reports*, 2(6):513–517, December 2000.

[4] M O A Sommer, G Dantas, and G M Church. Functional Characterization of the Antibiotic Resistance Reservoir in the Human Microflora. *Science (New York, N.Y.)*, 325(5944):1128–1131, August 2009.

[5] Abigail A Salyers, Anamika Gupta, and Yanping Wang. Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends in microbiology*, 12(9):412–416, September 2004.

[6] P Bidet, B Burghoffer, V Gautier, N Brahimi, P Mariani-Kurkdjian, A El-Ghoneimi, E Bingen, and G Arlet. In Vivo Transfer of Plasmid-Encoded ACC-1 AmpC from Klebsiella pneumoniae to Escherichia coli in an Infant and Selection of Impermeability to Imipenem in K. pneumoniae. *Antimicrobial agents and chemotherapy*, 49(8):3562–3565, July 2005.

[7] N Karami, A Martner, V I Enne, S Swerkersson, I Adlerberth, and A E Wold. Transfer of an ampicillin resistance gene between two Escherichia coli strains in the bowel microbiota of an infant treated with antibiotics. *Journal of Antimicrobial Chemotherapy*, 60(5):1142–1145, September 2007.

[8] Michael J Morowitz, Vincent J Denef, Elizabeth K Costello, Brian C Thomas, Valeriy Poroyko, David A Relman, and Jillian F Banfield. Strain-resolved community genomic analysis of gut microbial colonization in a premature infant. *Proceedings of the National Academy of Sciences*, 108(3):1128–1133, January 2011.

[9] I Sharon, M J Morowitz, B C Thomas, E K Costello, D A Relman, and J F Banfield. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Research*, 23(1):111–120, January 2013.

[10] Christopher T Brown, Itai Sharon, Brian C Thomas, Cindy J Castelle, Michael J Morowitz, and Jillian F Banfield. Genome resolved analysis of a premature infant gut microbial community reveals a Varibaculum cambriense genome and a shift towards fermentation-based metabolism during the third week of life. *Microbiome*, 1(1):30, 2013.

[11] Forough Nowrouzian, Bill Hesselmar, Robert Saalman, Inga-Lisa Strannegård, Nils Åberg, Agnes E Wold, and Ingegerd Adlerberth. Escherichia coli in Infants'

Intestinal Microflora: Colonization Rate, Strain Turnover, and Virulence Gene Carriage. *Pediatric Research*, 54(1):8–14, July 2003.

[12] A Carattoli, E Zankari, A Garcia-Fernandez, M Voldby Larsen, O Lund, L Villa, F Moller Aarestrup, and H Hasman. In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrobial agents and chemotherapy*, 58(7):3895–3903, June 2014.

[13] L Villa, A Garcia-Fernandez, D Fortini, and A Carattoli. Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. *Journal of Antimicrobial Chemotherapy*, 65(12):2518–2529, November 2010.

[14] M A Mulvey, J D Schilling, and S J Hultgren. Establishment of a persistent Escherichia coli reservoir during the acute phase of a bladder infection. *Infection and immunity*, 69(7):4572–4579, July 2001.

[15] J D Schilling, M A Mulvey, C D Vincent, R G Lorenz, and S J Hultgren. Bacterial Invasion Augments Epithelial Cytokine Responses to Escherichia coli Through a Lipopolysaccharide-Dependent Mechanism. *The Journal of Immunology*, 166(2):1148–1155, January 2001.

[16] R Kotlowski, C N Bernstein, S Sepehri, and D O Krause. High prevalence of Escherichia coli belonging to the B2+D phylogenetic group in inflammatory bowel disease. *Gut*, 56(5):669–675, May 2007.

[17] D O Krause, A C Little, S E Dowd, and C N Bernstein. Complete Genome Sequence of Adherent Invasive Escherichia coli UM146 Isolated from Ileal Crohn's Disease Biopsy Tissue. *Journal of bacteriology*, 193(2):583–583, December 2010.

[18] Chitrita Debroy, Mandeep S Sidhu, Upal Sarker, Bhushan M Jayarao, Adam L Stell, Nathan P Bell, and Timothy J Johnson. Complete sequence of pEC14_114, a highly conserved IncFIB/FIIA plasmid associated with uropathogenic Escherichia coli cystitis strains. *Plasmid*, 63(1):53–60, January 2010.

[19] Dona Wijetunge, Kurundu Karunathilake, Atul Chaudhari, Robab Katani, Edward G Dudley, Vivek Kapur, Chitrita Debroy, and Subhashinie Kariyawasam. Complete nucleotide sequence of pRS218, a large virulence plasmid that augments pathogenic potential of meningitis-associated. *BMC Microbiology*, 14(1):203, August 2014.

[20] Swaine L Chen, Chia-Seui Hung, Jian Xu, Christopher S Reigstad, Vincent Magrini, Aniko Sabo, Darin Blasiar, Tamberlyn Bieri, Rekha R Meyer, and Philip Ozersky. Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a comparative genomics approach. *Proceedings of the National Academy of Sciences of the United States of America*, 103(15):5977–5982, 2006.

[21] C K Cusumano, C S Hung, S L Chen, and S J Hultgren. Virulence Plasmid Harbored by Uropathogenic Escherichia coli Functions in Acute Stages of Pathogenesis. *Infection and immunity*, 78(4):1457–1467, March 2010.

[22] D Smajs and G M Weinstock. The Iron- and Temperature-Regulated cjrBC Genes of Shigella and Enteroinvasive Escherichia coli Strains Code for Colicin Js Uptake. *Journal of bacteriology*, 183(13):3958–3966, July 2001.

[23] Ramy K Aziz, Daniela Bartels, Aaron A Best, Matthew DeJongh, Terrence Disz, Robert A Edwards, Kevin Formsma, Svetlana Gerdes, Elizabeth M Glass, Michael Kubal, Folker Meyer, Gary J Olsen, Robert Olson, Andrei L Osterman, Ross A Overbeek, Leslie K McNeil, Daniel Paarmann, Tobias Paczian, Bruce Parrello, Gordon D Pusch, Claudia Reich, Rick Stevens, Olga Vassieva, Veronika Vonstein, Andreas Wilke, and Olga Zagnitko. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC genomics*, 9(1):75, 2008.

[24] Anders Norman, Lars Hestbjerg Hansen, Qunxin She, and Søren Johannes Sørensen. Nucleotide sequence of pOLA52: A conjugative IncX1 plasmid from Escherichia coli which enables biofilm formation and multidrug efflux. *Plasmid*, 60(1):59–74, July 2008.

[25] Timothy J Johnson, Eliza M Bielak, Daniela Fortini, Lars Hestbjerg Hansen, Henrik Hasman, Chitrita Debroy, Lisa K Nolan, and Alessandra Carattoli. Expansion of the IncX plasmid family for improved identification and typing of novel plasmids in drug-resistant Enterobacteriaceae. *Plasmid*, 68(1):43–50, July 2012.

[26] Y Zhou, Y Liang, K H Lynch, J J Dennis, and D S Wishart. PHAST: A Fast Phage Search Tool. *Nucleic Acids Research*, 39(suppl):W347–W352, June 2011.

[27] S M Shepard, J L Danzeisen, R E Isaacson, T Seemann, M Achtman, and T J Johnson. Genome Sequences and Phylogenetic Analysis of K88- and F18-Positive

Porcine Enterotoxigenic Escherichia coli. *Journal of bacteriology*, 194(2):395–405, December 2011.

[28] Maria Smorawinska, Magdalena Szuplewska, Piotr Zaleski, Paweł Wawrzyniak, Anna Maj, Andrzej Plucienniczak, and Dariusz Bartosik. Mobilizable narrow host range plasmids as natural suicide vectors enabling horizontal gene transfer among distantly related bacterial species. *FEMS Microbiology Letters*, 326(1):76–82, November 2011.

[29] D R Zerbino and E Birney. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18(5):821–829, February 2008.

[30] Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4):357–359, April 2012.

[31] H Li, B Handsaker, A Wysoker, T Fennell, J Ruan, N Homer, G Marth, G Abecasis, R Durbin, and 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, August 2009.

[32] A R Quinlan and I M Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, March 2010.

[33] P J A Cock, T Antao, J T Chang, B A Chapman, C J Cox, A Dalke, I Friedberg, T Hamelryck, F Kauff, B Wilczynski, and M J L de Hoon. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423, May 2009.

[34] Stefan Kurtz, Adam Phillippy, Arthur L Delcher, Michael Smoot, Martin Shumway, Corina Antonescu, and Steven L Salzberg. Versatile and open software for comparing large genomes. *Genome Biology*, 5(2):R12, 2004.

[35] S F Altschul, W Gish, W Miller, E W Myers, and D J Lipman. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410, October 1990.

[36] Nabil-Fareed Alikhan, Nicola K Petty, Nouri L Ben Zakour, and Scott A Beatson. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC genomics*, 12(1):402, August 2011.

5.8 Supplementary Information

Supplementary Table 5.1| Overview of library preparation method and sequencing technology used for the isolates genome sequenced in this study.

Lineage	Sampling Time	Library Preparation	Sequencing Technology Output
Lineage A	2d	Nextera kit	PE
	9d	Shear+barcode	SE
	16d	Shear+barcode	SE
	32d	Shear+barcode	SE
	$2m_1$	Shear+barcode	SE
	$2m_2$	TruSeq kit	PE
	$6m_1$	Shear+barcode	SE
	$6m_2$	TruSeq kit	PE
	12m	Shear+barcode	SE
	UTI	Nextera kit	PE
Lineage B	9d	Shear+barcode	SE
	16d	Shear+barcode	SE
	32d	Nextera kit	PE

Supplementary Table 5.2| Table containing the SNPs from lineage B, including the annotation and whether the animo acid change was synonymous or non-synonymous. The isolates from days 2, 9 and 16 are left out as they did not have any SNP differences from the representative genome for the lineage (isolate taken at 2d).

Contig number	Position	2d	UTI	32d	$2m_1$	$2m_2$	$6m_1$	$6m_2$	12m	Annotation	Change
61	61701	A	G							6-phosphogluconate dehydrogenase, decarboxylating (EC 1.1.1.44)	nonsynonymous
56	6755	G		A	A	A	A	A	A		
5	176310	G			A	A	A	A	A	Dihydrofolate reductase (EC 1.5.1.3)	synonymous
93	189	G			A	A	A				
42	18614	G			A		A			Primosomal protein I	nonsynonymous
93	95	G			A		A				
93	104	G			A		A				
93	105	G			A		A				
93	171	G			A		A				
15	137066	G			A					Radical SAM family enzyme, similar to coproporphyrinogen III oxidase, oxygen-independent, clustered with nucleoside-triphosphatase RdgB	synonymous
24	77511	G			A					Transcriptional activator of maltose regulon, MalT	nonsynonymous
49	14209	G			A					FIG00613320: hypothetical protein	synonymous
4	85958	G			A					3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	synonymous
62	46358	G			A						
5	175792	T			C	C	C	C	C	Dihydrofolate reductase (EC 1.5.1.3)	nonsynonymous
20	275987	T			C		C			Glutamate synthase [NADPH] large chain (EC 1.4.1.13)	nonsynonymous
20	297334	T			C					FUSARIC ACID RESISTANCE PROTEIN FUSB / FUSARIC ACID RESISTANCE PROTEIN FUSC	nonsynonymous
21	19073	T			C					IS, phage, Tn	nonsynonymous
24	25690	T			C					Nitrite reductase [NAD(P)H] small subunit (EC 1.7.1.4)	nonsynonymous
34	44455	A			G					LSU ribosomal protein L19p	synonymous
70	8108	A			G					Putative Ton-B dependent hemine receptor	nonsynonymous
62	60490	C			T	T	T	T	T	Alpha-2-macroglobulin	synonymous
93	150	C			T		T				
10	126386	C			T					Ferredoxin-type protein NapG (periplasmic nitrate reductase)	nonsynonymous

Contig number	pos	2d	UTI	32d	$2m_1$	$2m_2$	$6m_1$	$6m_2$	12m	annotation	change
14	3707	C			T					Putative GTP-binding protein	nonsynonymous
20	214976	C			T					LppC putative lipoprotein	nonsynonymous
23	43422	C			T					Putative phosphatase YieH	nonsynonymous
16	43695	G				A				Inner membrane protein YiaA	nonsynonymous
20	74596	G				A				Hexuronate utilization operon transcriptional repressor ExuR	nonsynonymous
42	9944	G				A				putative tail protein	nonsynonymous
46	5067	G				A				Putative stability/partitioning protein encoded within prophage CP-933T	synonymous
5	123037	G				A				Cell division protein FtsQ	nonsynonymous
17	98463	T				C				Holliday junction DNA helicase RuvB	synonymous
24	60742	T				C				Transcription elongation factor GreB	nonsynonymous
30	120639	G				C				expressed protein	nonsynonymous
10	92885	A				G				Putative membrane protein	nonsynonymous
26	193630	A				G				Putative inner membrane protein	synonymous
26	218585	A				G					
30	97163	A				G				Glutamate Aspartate transport ATP-binding protein GltL (TC 3.A.1.3.4)	nonsynonymous
83	6188	A				G				FIG01045219: hypothetical protein	synonymous
26	171451	G					A			Inner membrane ABC transporter permease protein YcjP	stop
3	137385	T					C			FIG138056: a glutathione-dependent thiol reductase	synonymous
27	15149	C					T			IncF plasmid conjugative transfer pilus assembly protein TraU	nonsynonymous
51	40378	C					T			Chitinase (EC 3.2.1.14)	synonymous
10	196403	C						A	A	DedA family inner membrane protein YohD	nonsynonymous
15	10933	G						A	A	Protease III precursor (EC 3.4.24.55)	synonymous
16	34911	G						A	A	Outer membrane protein A precursor	nonsynonymous
20	182830	G						A	A	Inner membrane protein	synonymous
23	37891	G						A	A	6-phospho-beta-glucosidase (EC 3.2.1.86)	nonsynonymous
25	57550	G						A	A	ATP-dependent hsl protease ATP-binding subunit HslU	nonsynonymous
35	4250	G						A	A	O-antigen ligase	synonymous
39	14389	G						A	A	Ubiquinone biosynthesis monooxygenase UbiB	nonsynonymous
3	26626	G						A	A	Enoyl-CoA hydratase (EC 4.2.1.17) / 3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) / 3-hydroxybutyryl-CoA epimerase (EC 5.1.2.3)	nonsynonymous
4	39270	G						A	A	Arginine/ornithine antiporter ArcD	nonsynonymous
4	66817	G						A	A	Uncharacterized protein YtfN	stop
5	5863	G						A	A	Prolyl-tRNA synthetase (EC 6.1.1.15), bacterial type	nonsynonymous

Contig number	Position	2d	UTI	32d	$2m_1$	$2m_2$	$6m_1$	$6m_2$	12m	Annotation	Change
10	145962	G						A		Putative ABC transporter ATP-binding protein	nonsynonymous
15	147808	G						A		Ornithine decarboxylase (EC 4.1.1.17)	nonsynonymous
17	309028	T						A		FIG00638170: hypothetical protein	nonsynonymous
18	136370	G						A		Hydroxylamine reductase (EC 1.7.-.-)	synonymous
19	96733	G						A		Per-activated serine protease autotransporter enterotoxin EspC	nonsynonymous
1	53133	G						A		PTS system, sorbose-specific IIC component (EC 2.7.1.69)	nonsynonymous
4	95471	G						A			
53	22400	G						A		Single-stranded DNA-binding protein	synonymous
5	191026	G						A		Carnitine operon protein CaiE	nonsynonymous
147	554	T						C	C	hypothetical protein	nonsynonymous
20	183063	T						C	C	Inner membrane protein	nonsynonymous
3	70340	T						C	C		
3	122654	T						C	C	NADP-dependent malic enzyme (EC 1.1.1.40)	synonymous
43	100362	T						C	C	Flagellar hook-associated protein FlgK	nonsynonymous
49	16016	T						C	C	LysR family transcriptional regulator YafC	nonsynonymous
5	120353	T						C	C	Cell division protein FtsZ (EC 3.4.24.-)	nonsynonymous
5	170421	T						C	C	Outer membrane protein Imp, required for envelope biogenesis / Organic solvent tolerance protein precursor	nonsynonymous
62	31047	T						C	C	Uncharacterized protein YphG, TPR-domain containing	nonsynonymous
94	267	T						C	C		
13	17707	T						C		Phage head-to-tail joining protein	nonsynonymous
25	1495	T						C		Outer membrane vitamin B12 receptor BtuB	synonymous
53	13735	T						C		Holin-like protein CidA	synonymous
10	178934	A						G	G	GTP cyclohydrolase I (EC 3.5.4.16) type 1	nonsynonymous
21	56232	A						G	G	adherence and invasion outermembrane protein (Inv,enhances Peyer's patches colonization)	nonsynonymous
83	198	A						G	G		
35	2808	A						G			
43	106731	A						G		Ribosomal large subunit pseudouridine synthase C (EC 4.2.1.70)	nonsynonymous
5	62832	A						G		glutamyl-Q-tRNA synthetase	synonymous
77	4289	A						G		hypothetical protein	nonsynonymous
147	436	C						T	T		
17	79481	C						T	T	Flagellar biosynthesis protein FlhA	synonymous
17	216927	C						T	T	PTS system, chitobiose-specific IIC component (EC 2.7.1.69)	nonsynonymous
17	299242	C						T	T		

Contig number	Position	2d	UTI 32d	$2m_1$	$2m_2$	$6m_1$	$6m_2$	12m	Annotation	Change
30	65147	C					T	T	Ribonuclease I precursor (EC 3.1.27.6)	nonsynonymous
43	39245	C					T	T		
17	232929	C					T		Threonyl-tRNA synthetase (EC 6.1.1.3)	synonymous
26	200370	C					T			
3	94889	C					T		Sulfate and thiosulfate import ATP-binding protein CysA (EC 3.6.3.25)	synonymous
43	29182	C					T			
5	225896	C					T		Aspartokinase (EC 2.7.2.4) / Homoserine dehydrogenase (EC 1.1.1.3)	synonymous
7	65266	C					T		ATP-dependent Clp protease proteolytic subunit (EC 3.4.21.92)	synonymous
10	14792	G						A	Phosphate acetyltransferase (EC 2.3.1.8)	synonymous
122	2192	G						A	Pyruvate formate-lyase (EC 2.3.1.54)	nonsynonymous
15	104208	G						A	Protein involved in stability of MscS mechanosensitive channel	nonsynonymous
15	115558	G						A	PTS system, mannitol-specific IIB component (EC 2.7.1.69) / PTS system, mannitol-specific IIC component (EC 2.7.1.69)	nonsynonymous
17	296387	G						A	Purine nucleotide synthesis repressor	synonymous
18	176760	G						A	Dipeptide transport system permease protein DppC (TC 3.A.1.5.2)	nonsynonymous
18	200735	T						A	Glutamate transport membrane-spanning protein	nonsynonymous
19	113564	G						A	Flagellar biosynthesis protein FlhA	synonymous
20	29027	G						A	Malate synthase G (EC 2.3.3.9)	synonymous
21	38804	G						A	iron aquisition yersiniabactin synthesis enzyme (Irp2)	nonsynonymous
23	7535	G						A	Ribose ABC transport system, ATP-binding protein RbsA (TC 3.A.1.2.1)	synonymous
23	14472	G						A		
23	135025	G						A	3-deoxy-D-manno-octulosonic-acid transferase (EC 2.-.-.-)	nonsynonymous
24	31001	G						A	FIG006427: Putative transport system permease protein	synonymous
25	1435	G						A	Outer membrane vitamin B12 receptor BtuB	synonymous
26	254547	T						A	hypothetical protein	nonsynonymous
26	254550	T						A	hypothetical protein	nonsynonymous
30	39378	G						A	Enterobactin synthetase component F, serine activating enzyme (EC 2.7.7.-)	nonsynonymous
32	23323	G						A		
32	70369	G						A	PhnJ protein	synonymous
42	3849	G						A	12	nonsynonymous
51	23078	G						A	LSU ribosomal protein L4p (L1e)	synonymous

Contig number	Position	2d	UTI	32d	$2m_1$	$2m_2$	$6m_1$	$6m_2$	12m	Annotation	Change
53	1415	G							A	Glutamate-aspartate carrier protein	synonymous
53	6083	G							A	NrfC protein	synonymous
5	25753	G							A	Intramembrane protease RasP/YluC, implicated in cell division based on FtsL cleavage	nonsynonymous
5	103289	G							A	VgrG protein	nonsynonymous
5	296526	G							A	Putative inner membrane protein	synonymous
7	15002	G							A	Glucokinase, ROK family (EC 2.7.1.2)	synonymous
7	35837	G							A	Nucleoside-specific channel-forming protein Tsx precursor	synonymous
10	41534	T							C	FIG00638014: hypothetical protein	synonymous
10	155429	T							C	Putative metal chaperone, involved in Zn homeostasis, GTPase of COG0523 family	nonsynonymous
15	25680	T							C	Lysophospholipid transporter LplT	nonsynonymous
15	59616	T							C	Predicted oxidoreductase, Fe-S subunit	nonsynonymous
15	91830	T							C	2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase (EC 1.14.13.-)	synonymous
15	102370	T							C		
17	149347	T							C	FIG00639292: hypothetical protein	nonsynonymous
17	257077	T							C	Acyl-CoA dehydrogenase (EC 1.3.99.3)	nonsynonymous
19	125833	T							C	FIG00641604: hypothetical protein	synonymous
20	29055	T							C	Malate synthase G (EC 2.3.3.9)	nonsynonymous
20	166907	T							C	Integral membrane protein TerC	synonymous
20	260128	T							C	Uncharacterized ABC transporter, ATP-binding protein YrbF	nonsynonymous
21	13870	T							C	hypothetical protein	synonymous
24	56129	T							C	FIG00638818: hypothetical protein	nonsynonymous
24	106762	T							C	Aspartate-semialdehyde dehydrogenase (EC 1.2.1.11)	nonsynonymous
26	131738	T							C	DNA topoisomerase I (EC 5.99.1.2)	nonsynonymous
27	43829	T							C		
30	47089	T							C	Isochorismate synthase (EC 5.4.4.2) of siderophore biosynthesis	nonsynonymous
31	39231	T							C	Branched-chain amino acid ABC transporter, amino acid-binding protein (TC 3.A.1.4.1)	nonsynonymous
32	37751	T							C	Lysyl-tRNA synthetase (class II) (EC 6.1.1.6)	synonymous
39	73336	T							C	Arylsulfatase (EC 3.1.6.1)	nonsynonymous
39	109139	T							C	Protein yifE	synonymous
4	23661	T							C	FIG000988: Predicted permease	nonsynonymous
56	21039	T							C	ATPase provides energy for both assembly of type IV secretion complex and secretion of T-DNA complex (VirB11)	nonsynonymous
5	61314	T							C	Sugar/maltose fermentation stimulation protein homolog	synonymous

Contig number	Position	2d	UTI	32d	$2m_1$	$2m_2$	$6m_1$	$6m_2$	12m	Annotation	Change
5	85250	T							C	Blue copper oxidase CueO precursor	synonymous
5	127065	T							C	Cell division protein FtsW	nonsynonymous
5	278788	T							C	TRAP dicarboxylate transporter, DctM subunit, unknown substrate 8	nonsynonymous
5	324641	T							C	Phosphoenolpyruvate-dihydroxyacetone phosphotransferase (EC 2.7.1.121), subunit DhaM	nonsynonymous
61	31544	T							C	Propanediol dehydratase reactivation factor large subunit	nonsynonymous
61	51403	T							C		
7	53127	T							C	FIG01057005: hypothetical protein	nonsynonymous
10	106855	A							G	Sensory histidine kinase AtoS	nonsynonymous
11	4616	A							G	Single-stranded DNA-binding protein	nonsynonymous
14	17663	A							G	GTP-binding protein TypA/BipA	nonsynonymous
15	15860	A							G	FIG004819: Prepilin peptidase dependent protein B precursor	nonsynonymous
15	34683	A							G		
16	65330	A							G	L-xylulose/3-keto-L-gulonate kinase (EC 2.7.1.-)	nonsynonymous
17	102945	A							G	Cell wall endopeptidase, family M23/M37	nonsynonymous
17	117869	A							G	Putative amidohydrolase	synonymous
17	143930	A							G	Rtn protein	synonymous
17	164143	A							G	Putative uncharacterized protein YeaK	nonsynonymous
17	240500	A							G	Vitamin B12 ABC transporter, permease component BtuC	synonymous
18	227304	A							G	Putative membrane protein	synonymous
19	74778	A							G	oxidoreductase, aldo/keto reductase family	nonsynonymous
20	82805	A							G	Putative cell division protein precursor	synonymous
20	161873	A							G	Putative cytoplasmic protein	synonymous
24	48644	A							G	Multimodular transpeptidase-transglycosylase (EC 2.4.1.129) (EC 3.4.-.-)	synonymous
26	254549	A							G	hypothetical protein	synonymous
28	32170	A							G	BarA sensory histidine kinase (VarS GacS)	nonsynonymous
28	53920	A							G	Putative electron transfer flavoprotein subunit YgcQ	synonymous
28	87481	A							G	Formate hydrogenlyase transcriptional activator	nonsynonymous
30	59369	A							G	LysR-family transcriptional regulator YbeF	nonsynonymous
30	134664	A							G	SeqA protein, negative modulator of initiation of replication	nonsynonymous
31	21929	A							G	Nickel ABC transporter, periplasmic nickel-binding protein NikA (TC 3.A.1.5.3)	nonsynonymous

Contig number	Position	2d	UTI	32d	$2m_1$	$2m_2$	$6m_1$	$6m_2$	12m	Annotation	Change
32	75262	A							G	FIG00638130: hypothetical protein	nonsynonymous
37	18162	A							G	Protein yhjK	nonsynonymous
39	100861	A							G	Threonine dehydratase biosynthetic (EC 4.3.1.19)	nonsynonymous
3	158670	A							G	Exopolyphosphatase (EC 3.6.1.11)	nonsynonymous
43	66383	A							G	Major curlin subunit precursor CsgA	synonymous
4	86155	A							G	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	nonsynonymous
55	5078	A							G	Respiratory nitrate reductase beta chain (EC 1.7.99.4)	nonsynonymous
5	230361	A							G	Inner membrane protein CreD	nonsynonymous
62	46866	A							G	Iron-sulfur cluster regulator IscR	synonymous
62	59308	A							G	Alpha-2-macroglobulin	synonymous
7	2996	A							G	putative lipoprotein	nonsynonymous
7	55881	A							G	Cytochrome O ubiquinol oxidase subunit IV (EC 1.10.3.-)	nonsynonymous
9	29805	A							G	Protein ydgH precursor	synonymous
9	31986	A							G	NAD(P) transhydrogenase alpha subunit (EC 1.6.1.2)	synonymous
9	37980	C							G	Permeases of the major facilitator superfamily	nonsynonymous
10	192408	C							T	tRNA-dihydrouridine synthase C (EC 1.-.-.-)	nonsynonymous
10	228596	C							T	Scaffold protein for [4Fe-4S] cluster assembly ApbC, MRP-like	synonymous
10	232712	C							T	Fimbriae usher protein StcC	nonsynonymous
10	300700	C							T	Mannose-1-phosphate guanylyltransferase (GDP) (EC 2.7.7.22) / Mannose-6-phosphate isomerase (EC 5.3.1.8)	nonsynonymous

Supplementary Table 5.2| Table containing the SNPs from the A lineage, including the annotation and whether the animo acid change was synonymous or non-synonymous. The isolates from days 2, 9 and 16 are left out as they did not have any SNP differences from the representative genome for the lineage (isolate taken at 2d).

Position	pUTI89	lineage A	Original amino acid	Amino acid change	Annotation
17995	A	G	G		hypothetical protein
51718	G	T	D	Y	rsvB
53060	C	T			
68939	G	A	Q		hypothetical protein
69562	C	T	Q		hypothetical protein
80619	G	A	G	D	traE
91269	T	G	L		trbC

Supplementary Table 5.3| Replicon type results from PlasmidFinder, presence in the isolates and the corresponding plasmid. Subvariants of the replicon type from PlasmidFinder are given in parentheses.

Lineage	Isolates	Identified Replicon	% Identity	Associated Plasmid
A	all	Col156	98.70	pNK29-2
A	all	FII(29)	99.61	pNK29-2
A	all	FIB(AP001918)	96.63	pNK29-2
A	all	X1	98.66	pNK29
A	$6m_2$, 12m	Col(IMGS31)	100.00	pNK29-3
B	32d	X1	98.66	pNK29
B	all	p0111	98.87	not characterized
B	all	B/O/K/Z	93.29	not characterized
B	all	Col(MG828)	91.92	not characterized

Supplementary Table 5.4| Annotations for pNK29. Annotations were done using the RAST server, but modified to include homologous genes from pOLA52. Gene names in square brackets indicate the name of the homologue in pOLA52.

ORF No.	Start	Stop	Strand	Annotation
1	105	650	+	DNA distortion protein 1 [taxA]
2	653	1819	+	IncQ plasmid conjugative transfer DNA nicking endonuclease TraR (pTi VirD2 homolog) [taxC]
3	2172	2690	+	actX, homologous to actX on pOLA52
4	2620	2832	+	hypothetical protein
5	2865	3512	+	Peptidoglycan hydrolase VirB1, involved in T-DNA transfer [pilx1]
6	3490	3786	+	Major pilus subunit of type IV secretion complex, VirB2 [pilx2]
7	3808	6561	+	ATPase provides energy for both assembly of type IV secretion complex and secretion of T-DNA complex (VirB4) [pilx3-4]
8	6572	7330	+	Minor pilin of type IV secretion complex (VirB5) [pilx5]
9	7331	7618	+	IncQ plasmid conjugative transfer protein TraG
10	7627	8757	+	Integral inner membrane protein of type IV secretion complex (VirB6) [pilx6]
11	8885	9013	+	pilx7, homologous to pilx7 on pOLA52
12	9003	9716	+	Inner membrane protein forms channel for type IV secretion of T-DNA complex, VirB8 [pilx8]
13	9721	10650	+	Forms the bulk of type IV secretion complex that spans outer membrane and periplasm (VirB9) [pilx9]
14	10647	11852	+	Inner membrane protein of type IV secretion of T-DNA complex, TonB-like, VirB10 [pilx10]
15	11854	12885	+	ATPase provides energy for both assembly of type IV secretion complex and secretion of T-DNA complex (VirB11) [pilx11]
16	12888	14723	+	Type IV secretion system protein VirD4 [taxB]
17	14720	15133	+	hypothetical lipoprotein, homologous to pOLA52
18	15130	15432	+	IncN plasmid KikA protein
19	15533	16858	+	Phage tail fiber protein
20	16887	17114	+	hypothetical protein
21	17194	17712	+	putative nuclease
22	17794	19056	+	Cell division protein FtsH (EC 3.4.24.-);Ontology_term=KEGG_ENZYME:3.4.24.-
23	19060	19626	+	FIG01045518: hypothetical protein
24	19640	21793	+	DNA topoisomerase III (EC 5.99.1.2);Ontology_term=KEGG_ENZYME:5.99.1.2 (topB)
25	21790	21996	+	Haemolysin expression modulating protein
26	22012	22473	+	DNA-binding protein H-NS [hns]
27	22524	22787	+	hypothetical protein
28	22898	25699	-	Mobile element protein, homologous to tnpA of Tn3 pRPEC180_47
29	25670	25921	+	hypothetical protein
30	26067	26624	+	Mobile element protein, homologous to tnpR of pRPEC180_47 (transposon Tn3 resolvase)
31	26807	27667	+	Beta-lactamase (EC 3.5.2.6);Ontology_term=KEGG_ENZYME:3.5.2.6
32	27814	27948	-	hypothetical protein
33	28012	28128	+	hypothetical protein
34	28432	30354	-	L. lactis predicted coding region ORF00041
35	30398	30754	+	DNA distortion protein 3
36	30751	31728	-	hypothetical protein
37	31753	32034	-	hypothetical protein

ORF No.	Start	Stop	Strand	Annotation
38	32132	32794	-	Chromosome partitioning protein ParA [par]
39	33175	33819	+	Resolvase [res]
40	33935	34114	-	hypothetical protein
41	34145	34264	+	hypothetical protein
42	34687	34803	-	hypothetical protein
43	35175	35429	+	hypothetical protein
44	36056	36724	+	L. lactis predicted coding region ORF00041
45	36758	37204	-	bis, homologous to pOLA52 (FIG01048616: hypothetical protein)
46	37244	38080	-	pir, homologous to pOLA52
47	37673	38002	-	YagA protein
48	38095	38310	-	FIG01048886: hypothetical protein
49	38300	38545	-	FIG01047979: hypothetical protein
50	38590	38913	-	FIG01047054: hypothetical protein
51	39059	39340	-	RelE/StbE replicon stabilization toxin
52	39330	39581	-	RelB/StbD replicon stabilization protein (antitoxin to RelE/StbE)
53	39578	39751	-	hypothetical protein
54	40368	40499	-	hypothetical protein
55	40816	41652	+	PI protein
56	41692	42138	+	FIG01047678: hypothetical protein

Supplementary Table 5.5| Annotations for pNK29-3. Annotations were done using the RAST server, but modified to include homologous genes from pIGMS31.

ORF No.	Start	Stop	Strand	Annotation
1	324	923	+	hypothetical protein
2	972	2144	-	MOB mobilization protein, homologous to pIGMS31

CHAPTER 6

Cultivation-based multiplex phenotyping of the human gut microbiota allows targeted recovery of previously uncultured bacteria

Elizabeth A. Rettedal[1], Heidi Gumpert[1], Morten O. A. Sommer[1,2]*

1) Department of Systems Biology, Technical University of Denmark, DK-2800 Lyngby, Denmark.
2) Novo Nordisk Foundation Center for Biosustainability, DK-2900 Hørsholm, Denmark.

*Correspondence should be addressed to M.O.A.S. at msom@bio.dtu.dk.

Published in Nature Communications August 28th, 2014

6.1 Abstract

The human gut microbiota is linked to a variety of human health issues and implicated in antibiotic resistance gene dissemination. Most of these associations rely on culture-independent methods since it is commonly believed that gut microbiota cannot be easily or sufficiently cultured. Here we show that carefully designed conditions enable cultivation of a representative proportion of human gut bacteria, enabling rapid multiplex phenotypic profiling. We use this approach to determine the phylogenetic distribution of antibiotic tolerance phenotypes for 16 antibiotics in the human gut microbiota. Based on the phenotypic mapping, we tailor antibiotic combinations to specifically select for previously uncultivated bacteria. Utilizing this method we cultivate and sequence the genomes of 4 isolates, one of which apparently belongs to the genus *Oscillibacter*; uncultivated *Oscillibacter* strains have been previously found to be anti-correlated with Crohn's disease.

6.2 Introduction

The human gut microbiota is a diverse microbial environment consisting of mostly anaerobic bacteria containing up to 10^{12} cells per gram of fecal matter [1]. The links with human disease [2, 3, 4] and other health concerns such as obesity [5, 6] have made it an important area for research. Increasing concerns about antibiotic resistance have also led to further study in the human gut as it is an ideal environment for acquisition and transfer of resistance genes to occur [7, 8, 9, 10]. New advances in culture-independent methodologies to study microbial communities have revolutionized how we study the human microbiome [1, 11], but in the process use of culture-based approaches has declined. While culture-independent methods can address a range of complex questions [12, 13] these approaches cannot make the definitive links to bacterial physiology that culturing does. Consequently, there is an increasing interest in improving culturing techniques for studying the gut microbiota.

Advances in culturing techniques include encapsulation of bacteria into microdroplets [14, 15], diffusion chambers simulating the natural environment of the samples [16, 17], microfabricated cultivation chips [18], and design of more effective culture media [19, 20, 21] with high-throughput identification of the cultured bacteria [22]. Furthermore, there have been several studies looking at the effect of using alternative solidifying agents to agar [20, 21]. The use of combinations of media has also allowed

sampling of a more diverse collection of bacteria. Indeed, a recent study utilizing 212 different culturing conditions retrieved 340 bacterial species from 3 human stool samples [23].

In spite of recent advances it remains widely believed that our ability to culture the human gut microbiota is far from comprehensive and relies on complicated techniques that cannot be readily implemented in most laboratories. Consequently, there remains a need for relatively simple procedures to cultivate and characterize the human gut bacterial community.

Here we combine novel cultivation conditions with high-throughput sequencing to identify conditions that allow representative cultivation and multiplex phenotypic profiling of the gut microbiota, allowing cultivation of important previously uncultivated species.

6.3 Results

Anaerobic cultivation retrieves a substantial part of live cells

To identify optimal anaerobic cultivation conditions for cultivating the human gut microbiota we tested the growth of bacteria from fecal samples on 10 different solid media. These media were chosen to represent a range of nutrient concentrations and several had previously been shown to have good potential for recovery of gut bacteria [22, 24, 25]. The goal was to try to more closely represent the conditions in the gut so a careful consideration of media components was undertaken (Supplementary Discussion). The growth was monitored over 7 days and it was observed that the colony count leveled out by day 3 for the majority of media (Supplementary Fig. 6.1). One of the lower nutrient media had pronounced growth between days 5-7, suggesting continued growth with increased incubation times. The lower nutrient media (GM1 and GM2 (M9 agar and gellan)), GM3 (GMM agar), and GM4 (GMM gellan) media (see Supplementary Methods for complete media descriptions) also tended to have overall lower plate counts. About half of the tested media had roughly the same (10^{10}) final colony counts, although the growth patterns varied by medium. The GM7 (modified GAM agar) medium had the highest overall plate counts. Notably, a difference in solidifying agent (agar vs. gellan) affected the growth on all media (Supplementary Fig. 6.1).

It is commonly believed that only a minor part of the gut microbiota can be cultivated using any specific cultivation condition [26, 27]. To assess the extent to which

the colony counts correlated with the number of live cells present in our samples we determined the fraction of live cells by microscopy (Methods) in parallel to cultivation using three independent samplings of two unrelated human adult volunteers. By determining the viable cells counts over time we hoped to establish that similar counts could be repeated in the same individual and determine if there was a large variability between individuals. Notably, the percentage of culturable cells (plate counts/viable cell counts) had a modest variability over time and varied between 30-40% for the two volunteers (Fig. 6.1a). This demonstrates that a substantial fraction of the viable bacteria in the gut microbiota can form colonies under these conditions. In contrast, the viable cells and culturable cells only represented 44% and up to 17% of the total cells, indicating the presence of dead or non-viable cells in the sample.

Cultivation allows representative sampling of gut microbiota

To assess the diversity of microorganisms capable of growing under our different cultivation conditions relative to those identified using culture-independent methods we determined the phylogenetic diversity of bacteria grown on each of our different cultivation media. We first combined multiple dilutions from each media in equal proportions to better sample the microbes capable of growing on each media (Methods). Our goal was to identify bacteria that were culturable at some point during the 7-day growth period. We are confident that most colonies did not grow as co-cultures as visual observation showed most colonies did not appear to touch (Methods), so we don't anticipate that they represented a substantial proportion of the population. Colonies were resuspended from solid media and DNA was extracted. Previous work from Goodman et al. [22] indicated that inoculum from non-viable or non-growing cells does not represent more than 2% of sequences recovered from scraping plates (sample diluted 10,000-fold). Most cells we recovered were from samples diluted 100,000 to 1,000,000-fold, further decreasing the likelihood of DNA persisting through the incubation period from non-growing cells. The V6 region of the 16S rRNA was amplified and sequenced to yield phylogenetic profiles for each of the growth conditions. To remove the noise from errors in sequencing we applied a filtering step during the initial sequence processing as previously suggested for Illumina data [28] (Methods). After filtering, each unique sequence (100% similarity) was considered a phylogenetic tag and subsequent analyses are based on these tags.

A graph of the taxonomic classifications (order level) and tag counts of each of

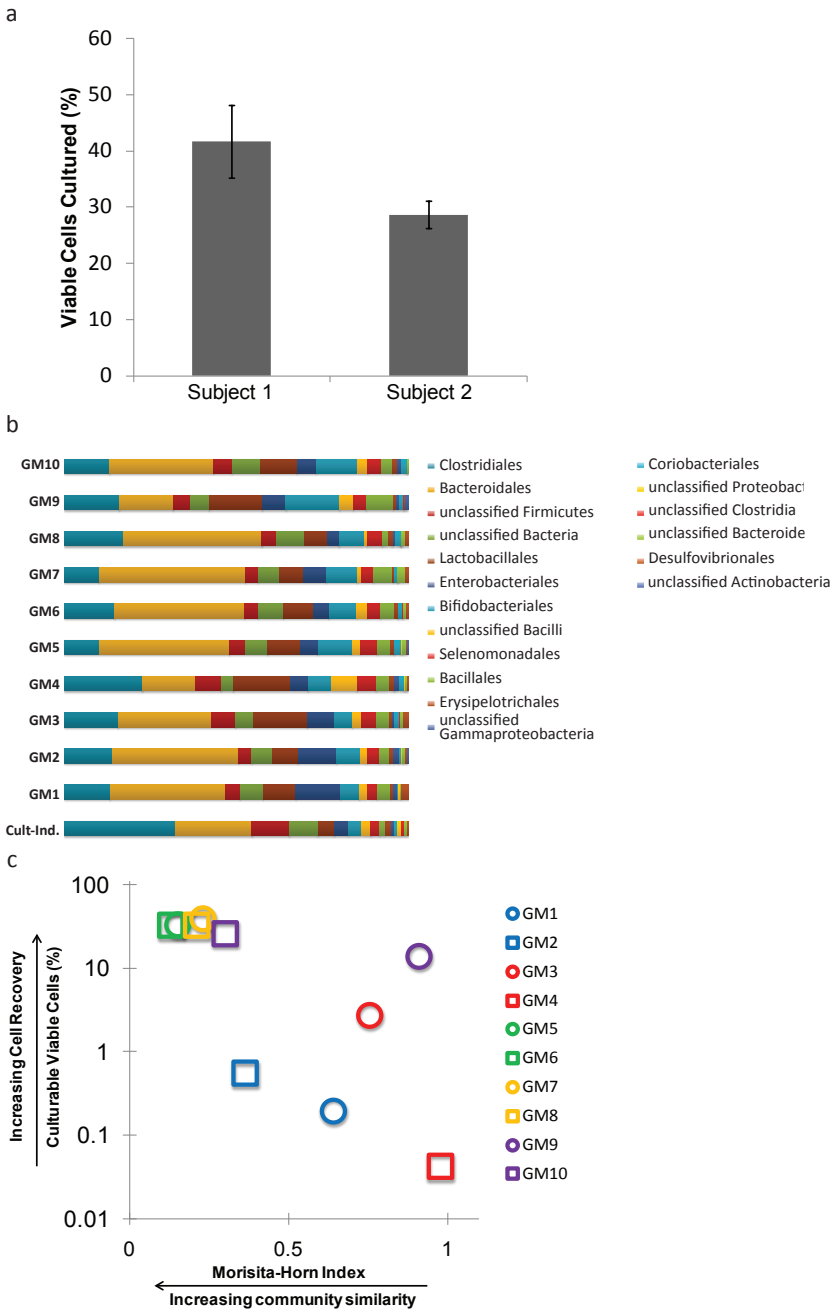


Figure 6.1 | Effectiveness of culturing conditions for recovery of human gut bacterial community. (a) The percentage of viable cells that we were able to culture. It was

Figure 6.1 continued | determined by dividing the GM7 plate counts by the trypan blue viable cell counts. The viable cells cultured were calculated for fecal samples obtained from two human subjects over 3 time points to show reproducibility across time and subjects. (b) An order-level taxonomic comparison of all tested media and the culture-independent (Cult-Ind.) sample based on tag abundance (Sample D1-T1). The order of the taxa is by abundance of the culture-independent sample. Bacterial taxa represented by less than 5 sequences for all samples are not displayed. (c) An assessment of each culture media's effectiveness for the recovery of the human gut bacterial community. The smaller the Morisita-Horn dissimilarity index (range from 0 to 1) and the larger the percentage of recovered viable cells the better the culture media represented the culture-independent community. The x-axis represents the Morisita-Horn index and the y-axis indicates the percentage of viable cells cultured on each media (Sample D1-T1).

the media and culture-independent sample highlights the differences in taxonomic representation between the samples. It was observed that 66-81% of order-level and 50-71% of family-level taxonomic groups from the total fecal sample were recovered on the tested media (Fig. 6.1b). If all media were combined a total of 95% of order-level and 88% of family-level taxonomic groups were recovered. It was also observed that changing a solidifying agent increased selection (relative sequence abundance) for certain taxa (Supplementary Fig. 6.2a). Lactobacillaceae increased in GM4 (gellan) compared to GM3 (agar) (Supplementary Fig. 6.2b) as opposed to an increase of Enterococcaceae in GM3 (agar) compared to GM4 (gellan) (Supplementary Fig. 6.2c). In other cases, the nutritional components (media composition) caused a selection as observed for Actinobacteria and Enterococcaceae, which are enriched in the GM9 (diluted modified GAM agar) and GM10 (diluted modified GAM gellan) as compared to most other media (Supplementary Fig. 6.2d).

Ninety-three of the tags (54.4% of unique tags) from the culture-independent sample were not represented in any culture media (Supplementary Data 1). However, these tags made up only a minority of the total sequences (15.8%) indicating many were of low abundance. None of these low abundance tags represented more than 2% of the uncultivated community. As total bacterial cells in the gut are known to reach 10^{12} CFU/g, dead cells that would be unable to be cultured likely represent some of the discrepancy between the methods. However, since distinct bacterial populations

are unlikely to be composed of a single cell we would expect that most bacteria have viable representatives within the sample. While the observance of tags unable to be recovered by culturing was anticipated, we were able to recover tags that represented 84% of sequences (Supplementary Data 2) in the gut microbiota (uncultivated sample). Notably, these tags represented 86% of sequences in the uncultivated sample. Recovery from individual media ranged from 68-78% (sequences). We also observed the recovery of 113 tags in culture that were not found in the culture-independent sample (Supplementary Data 3). The identity of these tags varied between media and originated from a variety of taxonomic groups. Many of these tags were selected for by multiple media and could be found in varying abundances. Not all of them appeared to be of very low abundance. Indeed, a number of Clostridia/*Clostridium* tags were among the culture-only isolates. Clostridia are well-known endospore formers that can be difficult to lyse in spore form [29]. These endospores may form during passage from the body, but sporulate into vegetative cells due to more favorable culture conditions making their recovery more likely. Other taxa poorly recovered in the total fecal sample include Staphylococcaceae and Enterococcaceae, whose thick gram-positive cell walls can also present a barrier to extraction [30, 31]. We cannot rule out the possibility that some variation arose due to variation in technical replicates and because slightly different extraction methods were used for the fecal samples vs. the cultured samples. The results highlight that while both culture-dependent and independent methods can broadly capture the same bacterial constituents, they also select for distinct subsets of the gut microbiota.

We then quantified the similarity between the tags obtained from different cultivation media to the community as determined using a culture-independent approach using the Morisita-Horn index, which measures the dissimilarity in community structure. We chose this comparative index because we are not only interested in maximizing the recovery of tags shared with the uncultivated sample, but also in preserving overall community structure. A similar community structure could indicate that growth conditions more closely represent what is found in the gut, better representing the microbiota's natural environment. We found that media GM5 (GAM agar), GM6 (GAM gellan), GM7 (modified GAM agar), and GM8 (modified GAM gellan) had average Morisita-Horn indices of 0.18 (Morisita-Horn index can range from 0 to 1), which demonstrates that the overall community architecture is well represented in these media. This is supported by biological replicates of the culture-independent sample showing Morisita-Horn indices of 0.12, while technical replicates had Morisita-

Horn indices of less than 0.04 indicating homology in PCR replicates. An optimal media would have a low Morisita-Horn index relative to the cultivation-independent community and a high percentage recovery of live cells. Notably, media GM5, GM6, GM7, GM8, and GM10 closely clustered to show good recovery while the remaining media were scattered along the graph highlighting the factors causing their divergence from the culture-independent sample (Fig. 6.1c, Supplementary Fig. 6.3). The GM4 media was clearly the worst media while several nutritionally similar media (GM5, GM6, GM7, and GM8) were the best. The GM7 had a slightly higher recovery of viable cells compared to the other similarly composed media (GM5, GM6, and GM8) while maintaining a high community structure similarity to the culture-independent sample. Further discussion on the cultivation and media is available in the Supplementary Discussion.

Since human gut microbiomes across individuals and geography are highly variable [32], the same effectiveness of these media may not be reflected across all samples tested. We suggest an application of this methodology to determine media appropriate to specific studies. To show that similar recovery could be repeated across several samples, we applied our GM7 media to compare culture-dependent and independent samples (Supplementary Fig. 6.4). Our results demonstrate that the recovery of taxonomic groups was repeatable both over time in the same subject and between multiple subjects.

Cultivation-based multiplex antibiotic tolerance phenotyping

Using high-throughput sequencing of the cultured community we were able to demonstrate that a good representation of the human gut microbiota can be cultivated on several of our media. The use of cultivation to survey a bacterial community provides us with the advantage of being able to link phenotypic characteristics to particular taxa which cannot be as definitively characterized by culture-independent methods. This method of phenotyping via culturing can be easily multiplexed to simultaneously look at specific phenotypic characteristics within a range of taxa within a given microbial community, which we term cultivation-based multiplex phenotyping.

To demonstrate this, we mapped the antibiotic tolerance phenotypes of the human gut microbiota, which demonstrates functionality (phenotypic tolerance) rather than only observed genotypes [33]. A fecal sample from a healthy volunteer that had been antibiotic free for the preceding year was tested against 16 different antibiotics (Sup-

plementary Table 1). The antibiotic concentrations targeted 2x the MICs (average) of commensal gut bacteria (mostly Bacteroidetes, Enterococcaceae, Enterobacteriaceae, and *Lactobacillus*) based on Eucast data (http://mic.eucast.org/Eucast2/). These antibiotic concentrations are not necessarily clinically relevant, but demonstrate how multiplex culture-based phenotyping can be applied to study microbial phenotypes.

Without antibiotic selection we recovered $1.7x10^{10}$ cfu/g, while the presence of antibiotics reduced the recovery to $5.1x10^{7}$- $1.5x10^{10}$ cfu/g, except for the chloramphenicol plates, which did not support growth (Supplementary Fig. 6.5). The antibiotics clindamycin and metronidazole had the lowest prevalence of tolerance phenotypes in the sampled gut microbiota varying between 0.3% and 1.1%. This is consistent with their common use to target anaerobic bacterial infections [34]. In contrast, gentamycin had the highest prevalence of tolerance phenotypes (91.4%), which is consistent with previous reports highlighting its reduced capability to inhibit growth of anaerobic and facultative anaerobic bacteria under anaerobic conditions [35, 36]. In general tolerance phenotypes towards beta-lactams (ampicillin, cefepime, cefotaxime, cephalexin, cefuroxime, dicloxacillin, and piperacillin) were high; on average 33% of the recovered taxa (tags) were able to grow in the presence of various beta-lactam antibiotics (calculated by determining (average counts on beta-lactam plates)/(control plate counts)).

To assess which tags were retrieved under different antibiotic selective pressures, the two dilutions above the countable plate (2 replicates per dilution) were collected and sequenced after 5 days of incubation at 37°C (see Methods). The tags were taxonomically classified (Supplementary Data 4) and a heat map was created in R to visualize the data (Fig. 6.2a). These mapped tags represent those that were also recovered on control plates (bacteria making up the majority of the gut community). This data provides a representative view of how antibiotic tolerance phenotypes are distributed across taxonomic groups in the microbiota (Fig. 6.2b-e). A mapping of all tags recovered as adjusted by plate count recovery (Supplementary Fig. 6.6) allows a more accurately proportional representation of the tags recovered on each antibiotic.

The distribution of the number of tolerance phenotypes linked to a specific tag varied widely. However, there were a large number of highly tolerant bacteria present. Indeed 31% of the tags were linked to tolerance phenotypes towards all tested antibiotics (Fig. 6.2b). The recoveries of tags broadly resistant to antibiotics are most likely associated with highly abundant tags that could comprise several bacterial species. The taxonomic classification of the most tolerant tags highlighted that Porphyromon-

adaceae, Bacteroidaceae, Enterobacteriaceae, and Enterococcaceae were the most tolerant taxa. Overall the Lachnospiraceae and Ruminococcaceae carried the least tolerance along with a number of unclassified Firmicutes and unknown Bacteria.

This phylogenetic mapping of the antibiotic tolerance phenotypes allows a better understanding of the tolerance levels of specific taxonomic groups in the gut microbiota. For instance many of our tolerant tags were classified as *Bacteroides* (Bacteroidetes) (Fig. 6.2c). Bacteria belonging to the phylum Bacteroidetes are among the most abundant in the human gut making up 20-30% of the total bacteria [9]. While there are some clinically relevant *Bacteroides* species (e.g. *Bacteroides fragilis*) they are generally considered commensals or mutualists [37]. Interestingly, *Bacteroides* have previously been associated with antibiotic resistance transfer events [9, 38]. In this study we find that 39% of the tags belonging to Bacteroidetes are associated with pan-tolerance phenotypes, highlighting their potential as a reservoir of antibiotic resistance.

Bacteria belonging to the phylum Proteobacteria, mostly Enterobacteriaceae, were also identified as highly tolerant in our sample. Enterobacteriaceae are of lower abundance in the gut, but the number of human pathogens and reports of resistance across classes of antibiotics associated with this group make it an important target of characterization [39, 40]. Consistent with these reports [8], we show that these tags are linked to tolerance towards most antibiotics tested (Fig. 6.2d).

We also recovered many tags on the antibiotic selection plates that were not found on the control plates (Supplementary Fig. 6.7 and 8). These likely represent bacterial community members of lower abundance (representing minorities of the community) that may have been missed by insufficient sequencing depth or sampling. However, antibiotics capable of reducing or suppressing growth of abundant strains could select for these less abundant populations provided they could grow in the presence of the antibiotics. Among those drugs that recovered more rare tags were those that reduced the growth of highly abundant Bacteroidaceae (*Bacteroides*) and/or Enterobacteriaceae populations (erythromycin, sulfamethoxazole, and ciprofloxacin) (Supplementary Fig. 6.7 and 9). These types of antibiotics could potentially be used to enrich cultivation for certain lower abundance populations of interest, as the medium itself was not the limiting factor in culturing these organisms. Indeed a previous study looking at the effect of ciprofloxacin treatment on human gut bacteria showed less perturbation in Lachnospiraceae and Clostridiales with treatment [41], supporting our enrichment for these groups on ciprofloxacin plates.

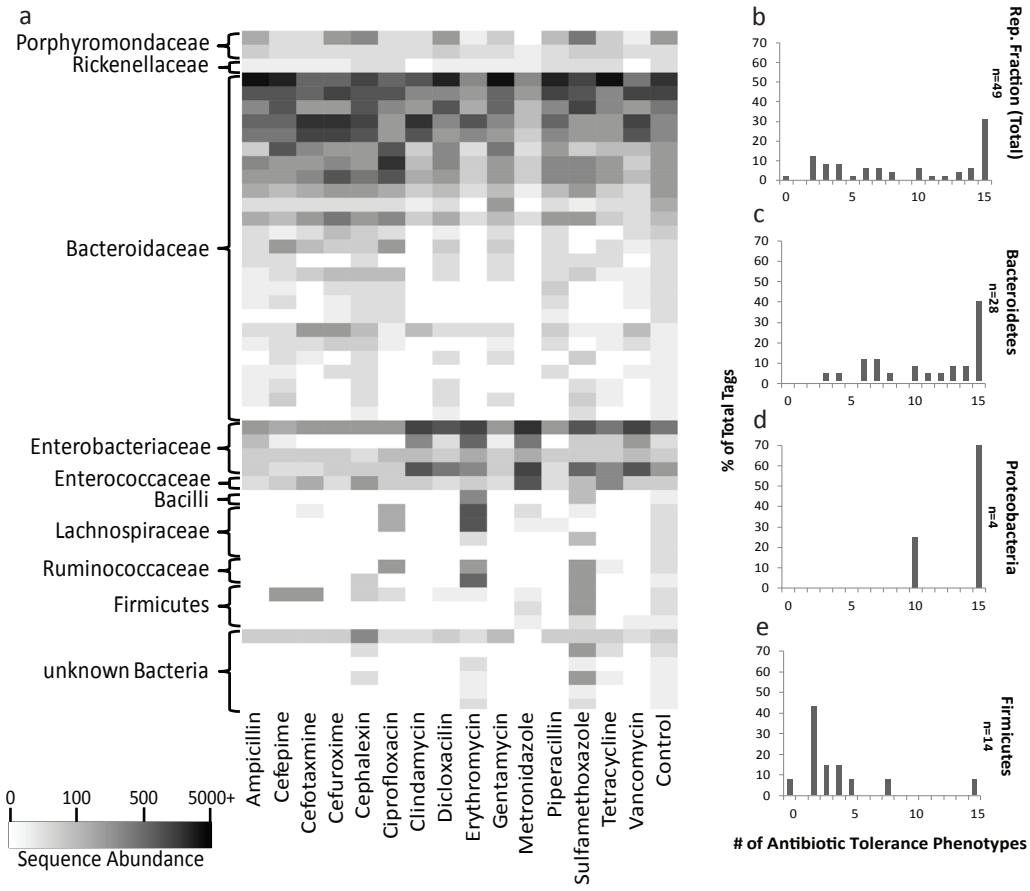


Figure 6.2 | Antibiotic tolerance distributions in majority gut bacterial populations. (a) The taxonomic classifications and proportions of tags representing the majority of cells within the gut community and their associated antibiotic tolerance phenotypes. The representative majority consists of the tags that were also recovered on the control plates. The tags are sorted by taxonomy and the highest level of classification was reported. (b) The number antibiotic tolerance phenotypes associated with each tag found in the representative majority of gut bacterial communities. (c-e) The number of tolerance phenotypes found per tag split into the (c) Bacteroidetes, (d) Proteobacteria, and (e) Firmicutes phylums. The number of tags associated with each phylum is also reported (n)=x. .

It should be noted that this technique is not able to distinguish between bacteria with intrinsic tolerance traits and transferable elements. Also, the association of a single tag with 15 tolerances doesn't mean that all bacteria with the same sequence carry all 15 tolerances. There are likely subpopulations of cells that only carry a fraction of the tolerance traits, but when looking at the total population of that particular tag at least one cell has tolerance to each of the antibiotics where tolerance was observed. Population-based resistance mechanisms, as demonstrated in E. coli communities [42], could also provide bacteria with protection from antibiotics without themselves carrying any tolerance traits.

Recovery of uncultivated bacteria using phenotypic mapping

To demonstrate the ability to enrich for subpopulations and recover specifically targeted taxa containing previously uncultured bacteria, a frozen sample of mixed bacteria preserved from antibiotic phenotyping plates (ciprofloxacin, sulfamethoxazole, and erythromycin) and a fresh sample from the same human subject were anaerobically cultured on antibiotic selective media (Methods). Using the Human Microbiome Project's (HMP) Most Wanted List $http://hmpdacc.org/most\_wanted/$ as a guide we tailored a combination of antibiotics based on the phenotyping data (Supplementary Fig. 6.9) to select for those taxa listed, specifically targeting *Odoribacter*, *Oscillibacter*, *Ruminococcus*, and *Faecalibacterium* (Fig. 6.3a). One hundred ninety-two individual colonies were screened via 16S Sanger sequencing to determine their identity. A total of 106 sequences were retrieved with 26 unique isolates identified. Five of those isolates had 95% or less identity to previously cultured isolates (Supplementary Data 5).

Genomic sequencing was performed on 4 of the 26 recovered isolates (2 of those with identities of less than 95% to a cultured isolate) and MIC testing was done to confirm the previously collected phenotyping data (Fig. 6.3b-e). The effectiveness of combining antibiotics to selectively culture isolates of interest can be clearly demonstrated with the increased recovery of the targeted genera (*Eubacterium* and *Blautia*) (Fig. 6.3c+d) or strains (P1C2 and P2C1) (Fig. 6.3b+e) when antibiotic combinations are used vs. single drugs. Genomic sequencing confirmed that two of these genomes had 16S rRNA identities of less than 95% to previously cultured bacteria suggesting they are likely to be new species. Isolates P1C2 and P1G4 had a high similarity (>99%) to previously cultured isolates *Intestinimonas butyriciproducens* and

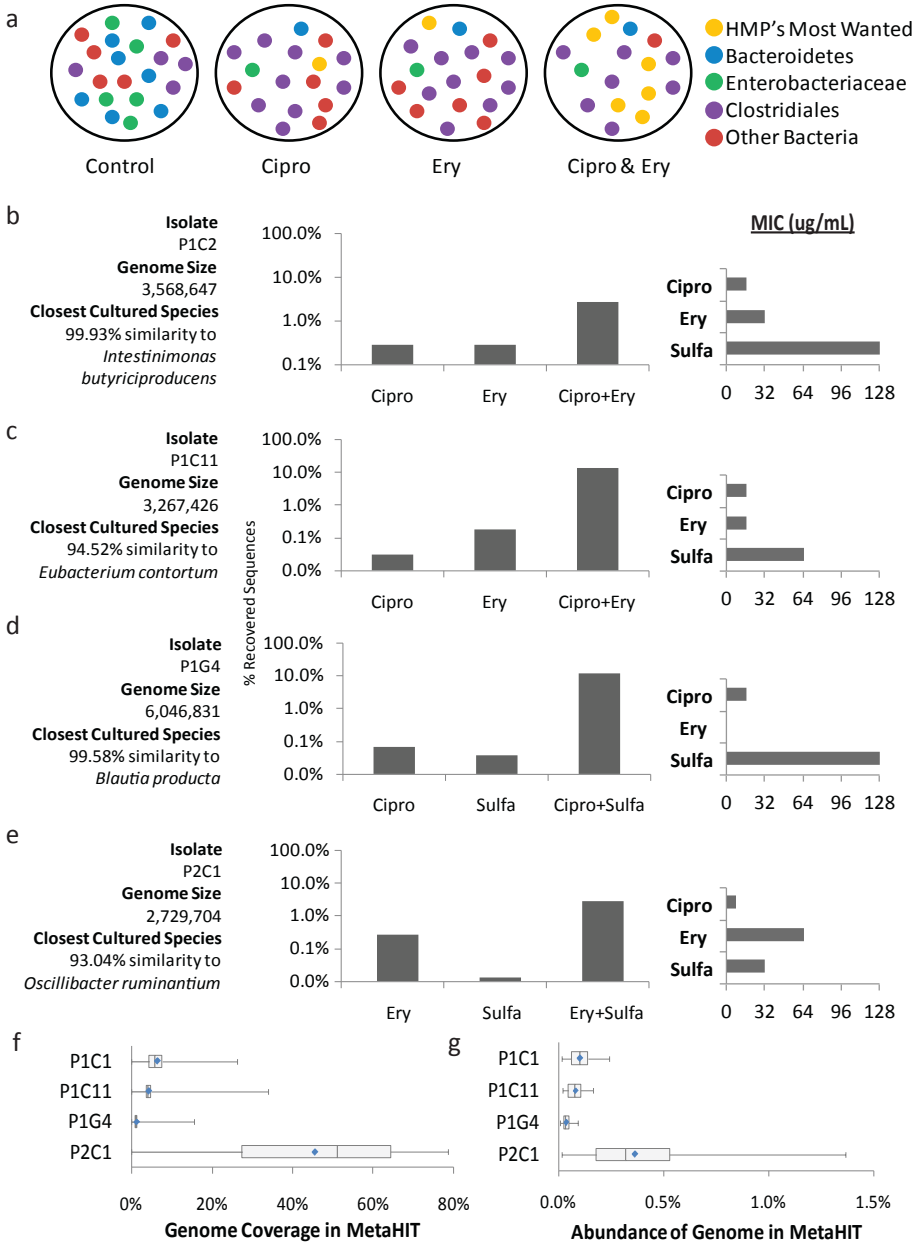


Figure 6.3 | Tailoring cultivation conditions allows recovery and characterization of specific bacteria. (a) The use of particular antibiotics (or combinations) helps select for the cultivation of bacteria of interest (previously uncultured and those on the

Figure 6.3 continued | Human Microbiome Project's (HMP) Most Wanted List). Particular antibiotics (e.g. Ciprofloxacin (Cip), Erythromycin (Ery), Sulfamethoxazole (Sulfa)) preferentially select for bacterial taxa that have few cultured isolates as compared with non-selection (Control). Combining antibiotics helps to further narrow the recoverable bacteria by exclusion of strains susceptible to both antibiotics. Therefore, particular combinations of drugs (e.g. Cip and Ery) are able to increase the recovery of strains that belong to poorly cultured taxa (such as HMP's Most Wanted List). (b-e) The 4 bacteria cultivated and genome sequenced based on multiplex phenotyping. The isolate name, genome size, and closest similarity (16S) to a previously cultured bacterium are indicated. A graph of the percentage of bacteria from either a specific isolate (b+e) or the genera Eubacterium (c) and Blautia (d) recovered on culture plates with a single antibiotic vs. a combination of antibiotics demonstrates successful targeted recovery based on phenotyping data. The MIC data on ciprofloxacin, erythromycin, and sulfamethoxazole tolerance for each of the 4 sequenced isolates is also reported. (f-g) The four genomes mapped to publically available MetaHIT metagenome data to determine their percentage genome coverage and abundance. For the box plots, the whisker bars span the minimum to maximum values while the box defines the 1st quartile, the median, and the 3rd quartile. The blue diamond denotes the mean. (f) The genome coverage represents the percentage of each of the 4 genomes that is covered by at least one read among MetaHIT metagenomes. (g) The abundance represents the total percentage of the metagenomes represented by each of the 4 sequenced genomes.

Blautia producta. Our novel isolates were most closely related to species *Eubacterium contortum* (94.8% similarity) (P1C11) and *Oscillibacter ruminantium* (93% similarity) (P2C1). The abundance of these isolates in the uncultivated library was very low or undetected (Supplementary Table 2). RAST [43] annotations indicated that all bacteria carried multiple antibiotic resistance genes conferring resistance to aminoglycosides, streptothricin, vancomycin, tetracycline, beta-lactams, fluoroquinolones, and multidrug efflux pumps.

To assess the abundance of these new isolates in gut metagenomes we mapped the reads from the publicly available MetaHIT [44] data to the sequenced genomes. We first determined what percentage of each of our genomes was covered in the MetaHIT samples and then quantified the overall abundance of the four genomes

within the MetaHIT metagenomes (Fig. 6.3f+g). Isolates P1C2, P1C11, and P1G4 had low genome coverage and abundance among the MetaHIT data which might suggest they are of low abundance or these particular isolates are not ubiquitous to the human microbiome. However our *Oscillibacter*-like isolate had a high percentage of coverage across the MetaHIT samples (43% average) and abundances of up to 1.36% in some metagenomes. Metagenomic data has suggested the existence of 19 species of *Oscillibacter* without a reference genome [45]. Genome comparisons in RAST of our *Oscillibacter*-like isolate (P2C1) to *Oscillibacter ruminatium* GH1, our *Intestinimonas butyriciproducens* isolate (P1C2), an unrelated gut microbe *Bifidobacterium animalis* subsp. lactis AD011, and a non-gut organism *Geobacillus thermodenitrificans* NG80-2 were undertaken to determine if there were any regions of high sequence homology between isolates (Supplementary Figure 10). Surprisingly the *I. butyriciproducens* isolate had more regions of high homology than the *O. ruminantium* GH1 genome, but we are unable to speculate on functional similarities as they were mainly hypothetical proteins. Uncultured *Oscillibacter* have been previously associated as responders to dietary change and as depleted in patients suffering from Crohn's disease [46, 47]. The ability to cultivate such organisms and study their physiology will contribute to our understanding of gastrointestinal disorders and devising treatment strategies.

6.4 Discussion

It is commonly believed that cultivation-based approaches are only able to capture a small part of organisms within the gut microbiota. Yet, our study has identified factors that can successfully help recover a representative fraction of the bacterial communities found in the human gut using techniques that can be readily implemented in most labs. Optimized cultivation procedures allow for multiplex mapping of phenotypes by directly coupling selective culture conditions with phylogenetic tag sequencing. In this way we surveyed the prevalence of antibiotic phenotypes in the human gut microbiota. The tolerances can be directly linked to specific taxa and provide useful phenotypic information that cannot readily be derived from culture-independent studies. We sampled a substantial proportion of the cells (10^{10}) in the human gut microbiota and were able to identify both common and rare tolerances. In this way we were able to detect a variety of taxonomic groups and assess the phylogenetic distribution of antibiotic tolerance phenotypes.

Cultivation-based multiplex phenotyping is a quick and easy way to screen for

antibiotic tolerance and look for tolerance patterns in human gut bacteria and should be a useful approach for studying the response to antibiotic treatment of the gut microbiome as well as illuminate possible risk factors predisposing individuals for antibiotic-associated diarrhea and other diseases.

Antibiotic tolerance phenotyping constitutes one useful application of cultivation-based multiplex phenotyping; however, this approach also has the potential to be used to identify substrate-specific utilization phenotypes as well as other tolerance phenotypes, including tolerance towards bile salts or high/low pH. This methodology could also be modified and applied to other environments such as soil or water to characterize phenotypic traits of interest. Notably, direct phenotyping can identify phenotypic characteristics, which cannot be as definitively characterized on a genetic basis and linked to specific bacteria using culture-independent methods.

Finally, mapping the antibiotic tolerance profiles among bacteria in the gut allowed us to target the recovery of specific taxa with previously uncultured isolates. As a proof of concept we identified 4 isolates, 2 of them novel species. One of these isolates appears to be a novel species of *Oscillibacter*; uncultivated strains belonging to this genus are significantly reduced in patients suffering from Crohn's disease. The ability to cultivate novel species with potential health benefits could be of interest to the probiotics industry and facilitate the development of novel probiotic supplements.

6.5 Methods

Sample Collection

All fecal samples were collected from 2 healthy adult human donors that had not taken antibiotics in the past year. One donor (D1) supplied three samples (T1, T2, T3), while the other donor (D2) supplied four samples over time (T1, T2, T3, T4). Sample D1-T1 was utilized for the initial media culture comparison, samples D1-T2 and all D2 samples were used to confirm reproducibility of culturing results, sample D1-T2 was used for the antibiotic tolerance phenotyping, and samples D1-T2 (frozen cultivated sample) and D1-T3 for directed recovery of bacteria. Consent was gained from all participants and proper ethical guidelines (Hvidovre Hospital) were followed throughout the study. Samples were transported to an anaerobic chamber within 5 minutes of collection. Five grams was separated out for culturing and 2.5 grams was collected for total DNA extractions. The remaining sample was stored at -80°C.

Media

The complete compositions of all media are listed in the Supplementary Methods. Cultivation, Viable Cell Counts, and DNA Extraction All manipulations were performed in a Type B Vinyl Anaerobic Chamber (Coy Lab Products) filled with an atmosphere of 95% nitrogen and 5% hydrogen.

Five grams of fecal sample was homogenously resuspended in 50 mL of 1X PBS that had been pre-reduced with resazurin (0.1 mg/mL). Ten-fold serial dilutions were performed to 10-9. Fifty µL of the 10-3, 10-4, and 10-5 dilutions were combined in equal proportion with 0.4% trypan blue. The countable dilution was identified and a hemocytometer and light microscope were used to perform a viable cell count in the anaerobic chamber. Viable cells counts were done in triplicate for each sample (per time point), and viable counts were performed three separate times for the same individual to make sure that recovery counts were consistent over time. The samples utilized for the viable cell counts included D1-T1, D1-T2, D1-T3, D2-T1, D2-T2, and D2-T3. Total cell counts were also performed on two samples from D1 to determine the fraction of cells that were not viable and/or culturable. We never observed the culturable cells as a higher number than the viable or total cells.

One hundred microliters of the serial dilutions were plated in 5 replicates on each of the 10 media (Supplementary Methods) to be tested. Seven to twelve sterile glass beads were used to spread the liquid on the plates through even shaking. Plates were anaerobically incubated at 37°C for 7 days. Plate counts were conducted each day to track the growth over time. On day 7, 3 dilutions, consisting of the countable dilution (between 25-250 colonies) and the dilutions above and below it, were chosen to select for a variety of plate crowding conditions (exceptions occurred for the Minimal and GMM media, where 4 dilutions were collected to accommodate lower overall plate counts). None of the plate dilutions utilized represented a situation where a lawn was formed or the plate would not be able to be counted. While it is not possible to confirm that all colonies scraped off the plates grew as individual colonies rather than co-cultures, the visual observations of the plates indicated that most colonies did not appear to touch. Further evidence to support this observation is that 95% of individual colonies picked into 96-well plates (see below) from dilution plates resulted in the identification of a single species. These dilutions were manually scraped off the media surface following a rinse with 5 mL 1x PBS and collected into 10 mL tubes. The collected tubes were centrifuged at 5000 x g for 30 minutes and remnants of solid

media were carefully removed from the pelleted samples. The pelleted samples were then extracted as individual dilutions (except Minimal and some GMM dilutions) using the MoBio UltraClean Microbial DNA Isolation kit.

The uncultivated total fecal sample was extracted using 2.5 grams of sample and the MoBio PowerMax Mega Soil DNA Isolation kit.

PCR and Sequencing

Extracted DNA was diluted to 5 ng/µL for all samples in order to keep quantities constant for the amplification step. PCR was performed using 1x Phusion High-Fidelity PCR Master Mix and tagged Illumina primers (10 µM concentration) in 30 µL reactions targeting the hypervariable V6 region (positions 969-1046) (primer and barcode sequences listed in Supplementary Data 6). Illumina overhangs (100 µM concentration) were attached in a second PCR reaction by combining barcoded samples in equal amounts as template for amplifying multiple 30 µL reactions. Thermocycling conditions for both PCR steps were as follows: initial denaturation 98°C for 1 min, followed by 30 cycles of 98°C for 10 sec, 65°C for 30 sec, and 72°C for 30 sec, and a final elongation at 72°C for 5 min. PCR product sizes were confirmed at each step and the final PCR product was gel purified using the Qiaex II Gel Extraction Kit (Qiagen). The samples were sequenced on an Illumina MiSeq.

Sequence Quality Control and Processing

Sequence quality control, processing, and analysis was performed using a modified version of the sequence quality control and analysis protocol [48] recommended for mothur (version 1.30.2) [49]. To pass the initial quality control, the sequence needed to (i) perfectly match the barcode, forward primer, and last 15 bases of the reverse primer, (ii) have a rolling window quality of 30 (rolling window size of 25); the sequences were truncated where they fell below quality window (iii) contain no ambiguous bases, (iv) have homopolymers of no more than 8 bases, and (v) have a minimal length of 60 bp [28, 48, 50]. Sequences that did not align along the same region (full V6) based on the Silva alignment were removed from the analysis. Sequences only occurring a single time (more than 50% of the unique sequences) were removed following the alignment step to denoise and reduce the dataset for further processing. The reverse primer and all but 4 bases of the forward primer were removed from the sequences. The 4 bases of the forward primer were kept as a reference to confirm

correct alignment throughout the analysis. A pre-clustering step in mothur (1 bp difference) was performed as recommended to reduce tag number inflation due to sequence errors [51]. Samples were checked for chimeras using UCHIME [52] and sequences identified as non-bacterial or unknown (80% cutoff) were removed from further analysis. Sequences were then compared at a 0% difference (unique) cutoff, giving each unique sequence its own taxonomic signature, which we refer to as a "tag". Sequences from multiple dilutions recovered from each media type were combined in equal proportions to better represent recovery from each media (e.g. For example, the sequences from the 3 GM5 dilutions (D5, D6, D7) were combined in equal proportion to represent the GM5 media). For each group of samples analyzed together, the sequences were sub-sampled (36040 for media comparison and 15450 for antibiotic phenotyping) to equalize the number of sequences per sample and rare sequence removal cutoff was established at 0.0166% abundance (abundance of 5 or less for 30000 sequences). Total tags and Morisita-Horn indices were calculated in mothur. BIOM files created in mothur were exported to MEGAN [53] to visualize taxonomic classifications on a family level and taxonomic classification and abundances as defined in mothur were used to construct further taxonomic graphs. Sequence abundance data and taxonomic classifications were used to create heat maps in R.

Culturing of Antibiotic Tolerant Bacterial Community

One hundred microliters of the serial dilutions were plated in 2 replicates on GM7 agar containing one of 16 antibiotics. The GM7 medium was chosen for the antibiotic tolerance testing because cultivation of samples on this medium showed good similarity to the uncultivated sample and because gellan-containing media has not yet been shown effective for antibiotic susceptibility testing. The selected antibiotic concentrations were based on Eucast data (http://mic.eucast.org/Eucast2/) of commensal gut bacteria, (mostly Bacteroidetes, Enterococcaceae, Enterobacteriaceae, and Lactobacillus), and were aimed for 2x higher than MICs (Supplementary Fig. 6.5). Seven to twelve glass beads were used to spread the liquid on the plates through even shaking. Plates were anaerobically incubated at 37°C for 5 days. These dilutions were manually scraped off the media surface following a rinse with 5 mL 1x PBS and collected into 10 mL tubes. One mL of the collected sample was centrifuged at 5000 x g for 30 minutes and remnants of solid media were carefully removed from the pelleted samples. DNA was then extracted from pelleted samples using the MoBio Ultra-

Clean Microbial DNA Isolation kit. The remaining 9 mL of the collected colonies was preserved at -80°C in 10% glycerol stocks.

Anaerobic Cultivation of New Bacterial Isolates

Frozen stocks of bacterial colonies (100 'l) previously scraped off ciprofloxacin, sulfamethoxazole, and erythromycin plates were plated in duplicate serial dilutions on GM7 media containing either erythromycin or sulfamethoxazole (concentrations as above). A fresh fecal sample from the same subject was used to make serial dilutions in duplicate on erythromycin and control (no antibiotics) plates. Plates were incubated anaerobically at 37°C for 5 days.

Ninety-six colonies from the frozen samples and 96 colonies from the fresh sample were freshly isolated on new media. Colony PCR was performed on the 16S rRNA region for the two 96-well plates using primers 8F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1391R (5'-GACGGGCGGTGTGTRCA-3') and Sanger sequenced using the forward primer. The sequences were quality control processed in CLC Main Workbench, and BLAST [54] and EzBioCloud [55] were used to identify the sequences. A total of 26 unique sequences were identified and the closest match to a cultured reference sequence was determined (Supplementary Data 5).

Genome Sequencing, Processing, and Analysis

Bacteria to be genome sequenced were anaerobically grown up in GAM broth for 1-3 days at 37°C. DNA from these bacteria was isolated using the MoBio UltraClean Microbial DNA Isolation Kit, genome sequence libraries were constructed using the Nextera DNA Sample Preparation Kit, and sequencing was performed on an Illumina MiSeq (150 bp paired-end reads). Genome assembly was done using Velvet (v1.09) [56], and contigs that were less than 500bp or with less than 3X coverage were filtered. The range of kmer values used were 93-125, and the assembly providing the best N50 score was used. Annotation of the assembled genomes was done using RAST [43]. RAST was also used to generate genome sequence comparisons between several isolates and publically available genomes.

Additionally, the high quality reads from the MetaHIT project [44] were aligned to each of the assembled genomes. Read alignment was done using Bowtie2 (v2.1.0) [57], with processing of the alignment files was done using SAMTools (v0.1.19) [58]. Coverage and sequencing depth of the aligned reads was assessed using BEDTools

(v2.17.0) [59]. The abundance of each assembled genome in the MetaHIT samples was calculated as the number of reads that aligned to the genome divided by the total number of reads.

6.6 Acknowledgements

This research was funded by the EU FP7-Health Program Evotar (282004) and The Lundbeck Foundation. M.O.A.S. acknowledges additional funding from the Novo Nordisk Foundation, Ib Henriksens Foundation and The Danish Free Research Councils.

Author Contributions

E.A.R. and M.O.A.S. conceived and designed the study. E.A.R. performed the experiments. H.G. assembled and analyzed the genomic data. E.A.R. and M.O.A.S. analyzed the data and prepared the manuscript.

Accession codes

Sequence reads for the 16S rRNA amplicon sequences have been deposited in the NCBI Sequence Read Archive under the BioSample code SAMN02869573, within BioProject PRJNA253252. The full 16S nucleotide sequences from the cultured isolates have been deposited in the GenBank NCBI database with accession codes KM043745 to KM043770. The genome sequences were deposited in the NCBI Whole Genome Shotgun database and can be found within BioProject PRJNA253252.

6.7 References

[1] E G Zoetendal, M Rajilic-Stojanovic, and W M de Vos. High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut*, 57(11):1605–1615, October 2008.

[2] Ilseung Cho and Martin J Blaser. The human microbiome: at the interface of health and disease. *Nature Reviews Genetics*, March 2012.

[3] Daniel N Frank, Allison L St Amand, Robert A Feldman, Edgar C Boedeker, Noam Harpaz, and Norman R Pace. Molecular-phylogenetic characterization of

microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 104(34):13780–13785, August 2007.

[4] Sharon Greenblum, Peter J Turnbaugh, and Elhanan Borenstein. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proceedings of the National Academy of Sciences of the United States of America*, 109(2):594–599, 2012.

[5] Peter J Turnbaugh, Micah Hamady, Tanya Yatsunenko, Brandi L Cantarel, Alexis Duncan, Ruth E Ley, Mitchell L Sogin, William J Jones, Bruce A Roe, Jason P Affourtit, Michael Egholm, Bernard Henrissat, Andrew C Heath, Rob Knight, and Jeffrey I Gordon. A core gut microbiome in obese and lean twins. *Nature*, 457(7228):480–484, January 2009.

[6] Peter J Turnbaugh, Ruth E Ley, Michael A Mahowald, Vincent Magrini, Elaine R Mardis, and Jeffrey I Gordon. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, 444(7122):1027–131, December 2006.

[7] Morten O A Sommer, George M Church, and Gautam Dantas. The human microbiome harbors a diverse reservoir of antibiotic resistance genes. *Virulence*, 1(4):299–303, July 2010.

[8] M O A Sommer, G Dantas, and G M Church. Functional Characterization of the Antibiotic Resistance Reservoir in the Human Microflora. *Science (New York, N.Y.)*, 325(5944):1128–1131, August 2009.

[9] Abigail A Salyers, Anamika Gupta, and Yanping Wang. Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends in microbiology*, 12(9):412–416, September 2004.

[10] Morten OA Sommer and Gautam Dantas. Antibiotics and the resistant microbiome. *Current Opinion in Microbiology*, 14(5):556–563, October 2011.

[11] Jean-Christophe Lagier, Matthieu Million, Perrine Hugon, Fabrice Armougom, and Didier Raoult. Human gut microbiota: repertoire and variations. *Frontiers in cellular and infection microbiology*, 2:136, 2012.

[12] Manimozhiyan Arumugam, Jeroen Raes, Eric Pelletier, Denis Le Paslier, Takuji Yamada, Daniel R Mende, Gabriel R Fernandes, Julien Tap, Thomas Bruls, Jean-Michel Batto, Marcelo Bertalan, Natalia Borruel, Francesc Casellas, Leyden Fernandez, Laurent Gautier, Torben Hansen, Masahira Hattori, Tetsuya Hayashi, Michiel Kleerebezem, Ken Kurokawa, Marion Leclerc, Florence Levenez, Chaysavanh Manichanh, H Bjørn Nielsen, Trine Nielsen, Nicolas Pons, Julie Poulain, Junjie Qin, Thomas Sicheritz-Ponten, Sebastian Tims, David Torrents, Edgardo Ugarte, Erwin G Zoetendal, Jun Wang, Francisco Guarner, Oluf Pedersen, Willem M de Vos, Søren Brunak, Joel Dore, María Antolín, François Artiguenave, Hervé M Blottiere, Mathieu Almeida, Christian Brechot, Carlos Cara, Christian Chervaux, Antonella Cultrone, Christine Delorme, Gérard Denariaz, Rozenn Dervyn, Konrad U Foerstner, Carsten Friss, Maarten van de Guchte, Eric Guedon, Florence Haimet, Wolfgang Huber, Johan van Hylckama-Vlieg, Alexandre Jamet, Catherine Juste, Ghalia Kaci, Jan Knol, Omar Lakhdari, Severine Layec, Karine Le Roux, Emmanuelle Maguin, Alexandre Mérieux, Raquel Melo Minardi, Christine M'rini, Jean Muller, Raish Oozeer, Julian Parkhill, Pierre Renault, Maria Rescigno, Nicolas Sanchez, Shinichi Sunagawa, Antonio Torrejon, Keith Turner, Gaetana Vandemeulebrouck, Encarna Varela, Yohanan Winogradsky, Georg Zeller, Jean Weissenbach, S Dusko Ehrlich, and Peer Bork. Enterotypes of the human gut microbiome. *Nature*, 473(7346):174–180, April 2011.

[13] Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402):207–214, April 2013.

[14] Karsten Zengler, Marion Walcher, Greg Clark, Imke Haller, Gerardo Toledo, Trevin Holland, Eric J Mathur, Gary Woodnutt, Jay M Short, and Martin Keller. High-Throughput Cultivation of Microorganisms Using Microcapsules. pages 124–130. Elsevier, 2005.

[15] Karsten Zengler, Gerardo Toledo, Michael Rappe, James Elkins, Eric J Mathur, Jay M Short, and Martin Keller. Cultivating the uncultured. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24):15681–15686, November 2002.

[16] A Bollmann, K Lewis, and S S Epstein. Incubation of Environmental Samples in

a Diffusion Chamber Increases the Diversity of Recovered Isolates. *Applied and Environmental Microbiology*, 73(20):6386–6390, October 2007.

[17] T Kaeberlein. Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment. *Science (New York, N.Y.)*, 296(5570):1127–1129, May 2002.

[18] Colin J Ingham, Ad Sprenkels, Johan Bomer, Douwe Molenaar, Albert van den Berg, Johan ET van Hylckama Vlieg, and Willem M de Vos. The micro-Petri dish, a million-well growth chip for the culture and high-throughput screening of microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*, 104(46):18217–18222, 2007.

[19] P H Janssen, P S Yates, B E Grinton, P M Taylor, and M Sait. Improved Culturability of Soil Bacteria and Isolation in Pure Culture of Novel Members of the Divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia. *Applied and Environmental Microbiology*, 68(5):2391–2396, May 2002.

[20] T Nyonyo, T Shinkai, A Tajima, and M Mitsumori. Effect of media composition, including gelling agents, on isolation of previously uncultured rumen bacteria. *Letters in Applied Microbiology*, 56(1):63–70, November 2012.

[21] Hideyuki Tamaki, Satoshi Hanada, Yuji Sekiguchi, Yasuhiro Tanaka, and Yoichi Kamagata. Effect of gelling agent on colony formation in solid cultivation of microbial community in lake sediment. *Environmental Microbiology*, 11(7):1827–1834, July 2009.

[22] Andrew L Goodman, George Kallstrom, Jeremiah J Faith, Alejandro Reyes, Aimee Moore, Gautam Dantas, and Jeffrey I Gordon. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proceedings of the National Academy of Sciences of the United States of America*, 108(15):6252–6257, 2011.

[23] J C Lagier, F Armougom, M Million, P Hugon, I Pagnier, C Robert, F Bittar, G Fournous, G Gimenez, M Maraninchi, J F Trape, E V Koonin, B La Scola, and D Raoult. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clinical Microbiology and Infection*, pages no–no, October 2012.

[24] Seong-Kyu Park, Min-Soo Kim, and Jin-Woo Bae. Blautia faecis sp. nov., isolated from human faeces. *International Journal of Systematic and Evolutionary Microbiology*, 63(Pt 2):599–603, 2013.

[25] Toshihiko Takada, Takashi Kurakawa, Hirokazu Tsuji, and Koji Nomoto. Fusicatenibacter saccharivorans gen. nov., sp. nov., isolated from human faeces. *International Journal of Systematic and Evolutionary Microbiology*, 63(Pt 10):3691–3696, 2013.

[26] Paul B Eckburg, Elisabeth M Bik, Charles N Bernstein, Elizabeth Purdom, Les Dethlefsen, Michael Sargent, Steven R Gill, Karen E Nelson, and David A Relman. Diversity of the human intestinal microbial flora. *Science (New York, N.Y.)*, 308(5728):1635–1638, June 2005.

[27] Mirjana Rajilić-Stojanović, Hauke Smidt, and Willem M de Vos. Diversity of the human gastrointestinal tract microbiota revisited. *Environmental Microbiology*, 9(9):2125–2136, September 2007.

[28] Nicholas A Bokulich, Sathish Subramanian, Jeremiah J Faith, Dirk Gevers, Jeffrey I Gordon, Rob Knight, David A Mills, and J Gregory Caporaso. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature methods*, 10(1):57–59, December 2012.

[29] CR Kuske, KL Banton, DL Adorada, PC Stark, KK Hill, and PJ Jackson. Small-Scale DNA Sample Preparation Method for Field PCR Detection of Microbial Cells and Spores in Soil. *Applied and Environmental Microbiology*, 64(7):2463–2472, July 1998.

[30] A Frostegård, S Courtois, V Ramisse, S Clerc, D Bernillon, F Le Gall, P Jeannin, X Nesme, and P Simonet. Quantification of bias related to the extraction of DNA directly from soils. *Applied and Environmental Microbiology*, 65(12):5409–5420, December 1999.

[31] I M Kauffmann, J Schmitt, and R D Schmid. DNA isolation from soil samples for cloning in different hosts. *Applied Microbiology and Biotechnology*, 64(5):665–670, June 2004.

[32] Tanya Yatsunenko, Federico E Rey, Mark J Manary, Indi Trehan, Maria Gloria Dominguez-Bello, Monica Contreras, Magda Magris, Glida Hidalgo, Robert N

Baldassano, Andrey P Anokhin, Andrew C Heath, Barbara Warner, Jens Reeder, Justin Kuczynski, J Gregory Caporaso, Catherine A Lozupone, Christian Lauber, Jose Carlos Clemente, Dan Knights, Rob Knight, and Jeffrey I Gordon. Human gut microbiome viewed across age and geography. *Nature*, May 2012.

[33] Gautam Dantas and Morten OA Sommer. Context matters — the complex interplay between resistome genotypes and resistance phenotypes. *Current Opinion in Microbiology*, 15(5):577–582, October 2012.

[34] J D Smilack, W R Wilson, and F R 3rd Cockerill. Tetracyclines, chloramphenicol, erythromycin, clindamycin, and metronidazole. *Mayo Clinic proceedings*, 66(12):1270–1280, December 1991.

[35] A V Reynolds, J M Hamilton-Miller, and W Brumfitt. Diminished effect of gentamicin under anaerobic or hypercapnic conditions. *Lancet*, 1(7957):447–449, February 1976.

[36] L E Bryan and S Kwan. Mechanisms of aminoglycoside resistance of anaerobic bacteria and facultative bacteria grown anaerobically. *The Journal of antimicrobial chemotherapy*, 8 Suppl D:1–8, December 1981.

[37] H M Wexler. Bacteroides: the Good, the Bad, and the Nitty-Gritty. *Clinical microbiology reviews*, 20(4):593–621, October 2007.

[38] N B Shoemaker, H Vlamakis, K Hayes, and A A Salyers. Evidence for Extensive Resistance Gene Transfer among Bacteroides spp. and among Bacteroides and Other Genera in the Human Colon. *Applied and Environmental Microbiology*, 67(2):561–568, February 2001.

[39] David L Paterson. Resistance in Gram-Negative Bacteria: Enterobacteriaceae. *The American Journal of Medicine*, 119(6):S20–S28, June 2006.

[40] Wolfgang Witte. Medical Consequences of Antibiotic Use in Agriculture. *Science (New York, N.Y.)*, 279(5353):996–997, February 1998.

[41] Les Dethlefsen, Sue Huse, Mitchell L Sogin, and David A Relman. The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. *PLoS biology*, 6(11):e280, 2008.

[42] Henry H Lee, Michael N Molla, Charles R Cantor, and James J Collins. Bacterial charity work leads to population-wide resistance. *Nature*, 467(7311):82–85, February 2010.

[43] Ramy K Aziz, Daniela Bartels, Aaron A Best, Matthew DeJongh, Terrence Disz, Robert A Edwards, Kevin Formsma, Svetlana Gerdes, Elizabeth M Glass, Michael Kubal, Folker Meyer, Gary J Olsen, Robert Olson, Andrei L Osterman, Ross A Overbeek, Leslie K McNeil, Daniel Paarmann, Tobias Paczian, Bruce Parrello, Gordon D Pusch, Claudia Reich, Rick Stevens, Olga Vassieva, Veronika Vonstein, Andreas Wilke, and Olga Zagnitko. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC genomics*, 9(1):75, 2008.

[44] Junjie Qin, Ruiqiang Li, Jeroen Raes, Manimozhiyan Arumugam, Kristoffer Solvsten Burgdorf, Chaysavanh Manichanh, Trine Nielsen, Nicolas Pons, Florence Levenez, Takuji Yamada, Daniel R Mende, Junhua Li, Junming Xu, Shaochuan Li, Dongfang Li, Jianjun Cao, Bo Wang, Huiqing Liang, Huisong Zheng, Yinlong Xie, Julien Tap, Patricia Lepage, Marcelo Bertalan, Jean-Michel Batto, Torben Hansen, Denis Le Paslier, Allan Linneberg, H Bjørn Nielsen, Eric Pelletier, Pierre Renault, Thomas Sicheritz-Ponten, Keith Turner, Hongmei Zhu, Chang Yu, Shengting Li, Min Jian, Yan Zhou, Yingrui Li, Xiuqing Zhang, Songgang Li, Nan Qin, Huanming Yang, Jian Wang, Søren Brunak, Joel Dore, Francisco Guarner, Karsten Kristiansen, Oluf Pedersen, Julian Parkhill, Jean Weissenbach, MetaHIT Consortium, Peer Bork, and S Dusko Ehrlich Jun Wang. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285):59–65, April 2010.

[45] Shinichi Sunagawa, Daniel R Mende, Georg Zeller, Fernando Izquierdo-Carrasco, Simon A Berger, Jens Roat Kultima, Luis Pedro Coelho, Manimozhiyan Arumugam, Julien Tap, Henrik Bjørn Nielsen, Simon Rasmussen, Søren Brunak, Oluf Pedersen, Francisco Guarner, Willem M de Vos, Jun Wang, Junhua Li, Joel Dore, S Dusko Ehrlich, Alexandros Stamatakis, and Peer Bork. Metagenomic species profiling using universal phylogenetic marker genes. *Nature methods*, 10(12):1196–1199, October 2013.

[46] Alan W Walker, Jennifer Ince, Sylvia H Duncan, Lucy M Webster, Grietje Holtrop, Xiaolei Ze, David Brown, Mark D Stares, Paul Scott, Aurore Bergerat, Petra Louis, Freda McIntosh, Alexandra M Johnstone, Gerald E Lobley,

Julian Parkhill, and Harry J Flint. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *The ISME Journal*, 5(2):220–230, August 2010.

[47] S Mondot, S Kang, J P Furet, D Aguirre de Carcer, C McSweeney, M Morrison, P Marteau, J Doré, and M Leclerc. Highlighting new phylogenetic specificities of Crohn's disease microbiota. *Inflammatory Bowel Diseases*, 17(1):185–192, January 2011.

[48] Patrick D Schloss, Dirk Gevers, and Sarah L Westcott. Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. *PloS one*, 6(12):e27310, December 2011.

[49] P D Schloss, S L Westcott, T Ryabin, J R Hall, M Hartmann, E B Hollister, R A Lesniewski, B B Oakley, D H Parks, C J Robinson, J W Sahl, B Stres, G G Thallinger, D J Van Horn, and C F Weber. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, 75(23):7537–7541, November 2009.

[50] Patrick D Schloss. The Effects of Alignment Quality, Distance Calculation Method, Sequence Filtering, and Region on the Analysis of 16S rRNA Gene-Based Studies. *PLoS Computational Biology*, 6(7):e1000844, July 2010.

[51] Susan M Huse, David Mark Welch, Hilary G Morrison, and Mitchell L Sogin. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology*, 12(7):1889–1898, March 2010.

[52] R C Edgar, B J Haas, J C Clemente, C Quince, and R Knight. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16):2194–2200, August 2011.

[53] Suparna Mitra, Mario Stärk, and Daniel H Huson. Analysis of 16S rRNA environmental sequences using MEGAN. *BMC genomics*, 12(Suppl 3):S17, November 2011.

[54] S F Altschul, W Gish, W Miller, E W Myers, and D J Lipman. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410, October 1990.

[55] O S Kim, Y J Cho, K Lee, S H Yoon, M Kim, H Na, S C Park, Y S Jeon, J H Lee, H Yi, S Won, and J Chun. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *International Journal of Systematic and Evolutionary Microbiology*, 62(Pt 3):716–721, March 2012.

[56] D R Zerbino and E Birney. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18(5):821–829, February 2008.

[57] Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4):357–359, April 2012.

[58] H Li, B Handsaker, A Wysoker, T Fennell, J Ruan, N Homer, G Marth, G Abecasis, R Durbin, and 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, August 2009.

[59] A R Quinlan and I M Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, March 2010.

6.8 Supplementary Discussion

Choosing an effective media for human gut bacteria

To choose cultivation media for this study, we utilized previous research to determine which media could effectively recover anaerobic human gut communities. These media represented different nutritional conditions and made use of multiple solidifying agents. Space is often a limitation in anaerobic studies, so high-throughput (short incubation times) potential was also considered, as it can be a key factor in determining the depth and breadth of studies.

Variations in the recovery of bacteria was observed across both different nutritional components and solidifying agents. While no single media recovered all the observed diversity, particular media better represented the bacterial community recovered from the uncultivated fecal sample.

Minimal Media (GM1 and 2)

In this study, the use of M9 minimal media (GM1 and 2) served as a baseline media with an expected low recovery as it is composed of only the essential components for the growth of bacteria with simple nutritional requirements. To try and encourage growth of slightly more fastidious organisms it was supplemented with additional vitamins and minerals, some of which would normally be available in the gut environment. As expected, lower plate counts and greater dissimilarity to the uncultivated sample was observed in this study.

GMM Media (GM3 and 4)

GMM (Gut Microbiota Medium) (GM3 and 4) is a recently devised media [1] that is a diluted form of tryptone-glucose-yeast supplemented (TGYS) media with modifications. It is a complex media containing multiple energy sources, including carboxylic/-fatty acids, vitamins, minerals, and supplements such as hematin and is formulated for the growth of gut microbes. Overall the GMM was not best performer in this study, particularly the gellan, as it vastly diverged in similarity to the uncultivated sample. The lower plate counts and domination by Lactobacillaceae suggested that gellan might have reacted with a media component to make it more selective. In the original study [1], GMM agar applied to human fecal samples in vitro recovered 70% of the genera found in an uncultivated sample. While the differences in recovery observed between that study and ours might be partially due to differences in the sample participants' microbial communities, we believe that much of the variation originates from how the sequences were clustered in the analysis. We utilized a 'tag' level clustering which required that the sequences have the same nucleotide sequence (100% similarity) to be considered present in compared samples, while in the Goodman et al. paper [1] they clustered at 97% and evaluated their recovery based on the sequences belonging to the same taxonomic classification (e.g. genera, family, ect.). We also evaluated our recovery based on more than a presence/absence measure by taking into account the structure of the recovered communities. Using their comparative methods, we observed that we could recover 81% of order-level and 71% of family-level taxonomic groups using GMM agar, which is more similar to their results.

GAM and MGAM Media (GM5, 6, 7, 8)

Gifu and modified Gifu anaerobe media (GAM and MGAM) (GM5, 6, 7, 8) are commercially available media principally used in anaerobic clinical and antibiotic susceptibility testing [2, 3]. The modified version contains diluted components and extra amino acids, vitamin K, and hemin, but both are complex media with a variety of energy sources. They have been used in a small number of studies [4, 5] and recently applied to recover novel bacteria from human fecal samples [6, 7, 8]. A large-scale study of their effectiveness had not been previously undertaken. In this study these media recovered the highest plate counts and had the best maintenance of community structure, suggesting its usefulness in cultivation of human fecal samples.

Diluted GAM Media (GM9 and 10)

Diluting media has been shown to select for previously uncultivated bacteria [9, 10] as it is thought that high nutrient media may exclude cultivation of many slower growing organisms [11]. Diluted versions of modified GAM (DMGAM) (GM9 and 10) with 1/10 the concentration of all nutrients was created to test this idea. Although the diluted version of the agar (GM9) showed less similarity to the uncultivated sample than the other GAM media, the DMGAM did recover more unique tags than the undiluted version (MGAM), supporting that lower nutrient media can help recovered greater diversity.

Solidifying Agents

While agar remains the standard solidifying agent for culture media, there has been increasing interest in alternatives such as gellan. Studies have shown that some bacteria recovered on gellan cannot be grown on agar even if the media's nutritional composition is the same [12, 13]. It has also been observed that gellan sometimes has the effect of accelerating colony formation [12]. However, we saw a more gradual increase in colony counts over time for gellan-containing media. A previous bacterial soil study found that gellan recovered higher numbers of viable cells, but did not change diversity [14]. We did not observe the same effect in our study in the human gut; rather we observed that the effect of gellan was media-dependent. Gellan-containing media in this study did not capture a greater number of tags overall, but they did tend to capture a different subset of tags than found in in agar-containing media. It is not yet completely understood why gellan can change bacterial recovery, but our

observations along with previous studies [12] suggest a media and/or environmental-dependent growth modulation. Gellan had not been previously utilized for human gut culturing, but this study has shown its effectiveness in this untested environment.

6.9 Supplementary References

[1] Goodman, A. L. et al. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. Proceedings of the National Academy of Sciences of the United States of America 108, 6252-6257, doi:10.1073/pnas.1102938108 (2011).

[2] Oka, K. et al. Molecular and microbiological characterization of Clostridium difficile isolates from single, relapse, and reinfection cases. J Clin Microbiol 50, 915-921, doi:10.1128/JCM.05588-11 (2012).

[3] Löfmark, S., Edlund, C. & Nord, C. E. Metronidazole Is Still the Drug of Choice for Treatment of Anaerobic Infections. Clinical Infectious Diseases 50, S16-S23, doi:10.1086/647939 (2010).

[4] Nagaoka, K. et al. Macrolides inhibit Fusobacterium nucleatum-induced MUC5AC production in human airway epithelial cells. Antimicrob Agents Chemother 57, 1844-1849, doi:10.1128/AAC.02466-12 (2013).

[5] Matsuda, K. et al. Establishment of an analytical system for the human fecal microbiota, based on reverse transcription-quantitative PCR targeting of multicopy rRNA molecules. Appl Environ Microbiol 75, 1961-1969, doi:10.1128/AEM.01843-08 (2009).

[6] Morotomi, M., Nagai, F. & Watanabe, Y. Description of Christensenella minuta gen. nov., sp. nov., isolated from human faeces, which forms a distinct branch in the order Clostridiales, and proposal of Christensenellaceae fam. nov. Int J Syst Evol Microbiol 62, 144-149, doi:10.1099/ijs.0.026989-0 (2012).

[7] Park, S. K., Kim, M. S. & Bae, J. W. Blautia faecis sp. nov., isolated from human faeces. Int J Syst Evol Microbiol 63, 599-603, doi:10.1099/ijs.0.036541-0 (2013).

[8] Takada, T., Kurakawa, T., Tsuji, H. & Nomoto, K. Characterization and classification of a novel microbe,Fusicatenibacter saccharivorans gen. nov., sp. nov., isolated from human feces. Int J Syst Evol Microbiol, doi:10.1099/ijs.0.045823-0 (2013).

[9] Rappe, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature 418, 630-633, doi:10.1038/nature00917 (2002).

[10] Zengler, K. et al. Cultivating the uncultured. Proc Natl Acad Sci U S A 99, 15681-15686, doi:10.1073/pnas.252630999 (2002).

[11] Koch, A. L. Microbial physiology and ecology of slow growth. Microbiol Mol Biol Rev 61, 305-318 (1997).

[12] Tamaki, H., Hanada, S., Sekiguchi, Y., Tanaka, Y. & Kamagata, Y. Effect of gelling agent on colony formation in solid cultivation of microbial community in lake sediment. Environ Microbiol 11, 1827-1834, doi:10.1111/j.1462-2920.2009.01907.x (2009).

[13] Nyonyo, T., Shinkai, T., Tajima, A. & Mitsumori, M. Effect of media composition, including gelling agents, on isolation of previously uncultured rumen bacteria. Letters in applied microbiology 56, 63-70, doi:10.1111/lam.12019 (2013).

[14] Janssen, P. H., Yates, P. S., Grinton, B. E., Taylor, P. M. & Sait, M. Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia. Applied and environmental microbiology 68, 2391-2396, doi:Doi 10.1128/Aem.68.5.2391-2396.2002 (2002).

6.10 Supplementary Methods

Media Recipes

Minimal Media (Supplemented) (1 L)

Disodium phosphate	6.78 g
Monopotassium phosphate	3 g
Sodium chloride	0.5 g
Ammonium chloride	1 g
ATCC Vitamin Mix	4 mL
ATCC Mineral Mix	4 mL
Sugar Mix (50 mM stock)	4 mL
-Arabinose L(+)	
-Galactose D(+)	
-Glucose D(+)	
-Maltose D(+)	
Vitamin K (1 mg/mL)	400 μL
Casamino Acids (20% stock)	4 mL (Final Concentration 0.1%)
Agar/Gellan	12 g

Gut Microbiota Media (adjust pH to ±7.0 before autoclaving) (1 L) (Goodman et al., 2011)

Component	Amount/L	Concentration	Comments
Tryptone Peptone	2 g	0.20%	
Yeast Extract	1 g	0.10%	
D-glucose	0.4 g	2.2 mM	
L-cysteine	0.5 g	3.2 mM	
Cellobiose	1 g	2.9 mM	
Maltose	1 g	2.8 mM	
Fructose	1 g	2.2 mM	
Meat Extract	5 g	0.50%	
KH2PO4	100 mL	100 mM	1M stock solution pH 7.2
MgSO4-7H20	0.002 g	0.008 mM	
NaHCO3	0.4 g	4.8 mM	
NaCl2	0.08 g	1.37 mM	
CaCl2	1 mL	0.80%	0.8g/100mL stock
Vitamin K (menadione)	1 mL	5.8 mM	1 mg/mL stock solution
FeSO4	1 mL	1.44 mM	0.4 mg FeSO4/mL stock solution
Histidine Hematin Solution	1 mL	0.10%	1.2 mg hematin/mL in 0.2M histidi
Tween 80	2 mL	0.05%	25% stock solution
ATCC Vitamin Mix	10 mL	1%	
ATCC Trace Mineral Mix	10 mL	1%	
Acetic acid	1.7 mL	30 mM	
Isovaleric acid	0.1 mL	1 mM	
Propionic acid	2 mL	8 mM	
Butyric acid	2 mL	4 mM	
Resazurin	4 mL	4 mM	0.25 mg/mL stock solution
Noble Agar/Gellan	12 g	1.20%	

GAM (Final pH ±7.1) (1 L)

Peptone	10.0 g
Soya Peptone	3.0 g
Proteose Peptone	10.0 g
Digested Serum	13.5 g
Yeast Extract	5.0 g
Meat Extract	2.2 g
Liver Extract	1.2 g
Dextrose	3.0 g
Potassium Dihydrogen Phosphate	2.5 g
Sodium Chloride	3.0 g
Soluble Starch	5.0 g
L-Cysteine Hydrochloride	0.3 g
Sodium Thioglycollate	0.3 g
Agar/Gellan	15 g

Modified GAM (Final pH ±7.3) (1 L)

Peptone	5.0 g
Soya Peptone	3.0 g
Proteose Peptone	5.0 g
Digested Serum	10.0 g
Yeast Extract	2.5 g
Meat Extract	2.2 g
Liver Extract	1.2 g
Dextrose	0.5 g
Soluble Starch	5.0 g
L-Tryptophan	0.2 g
L-Cysteine Hydrochloride	0.3 g
Sodium Thioglycollate	0.3 g
L-Arginine	1.0 g
Vitamin K1	5 mg
Hemin	10 mg
Potassium Dihydrogen Phosphate	2.5 g
Sodium Chloride	3.0 g
Agar/Gellan	15g

Diluted Modified GAM (Final pH ±7.3) (1 L)

Peptone	0.5 g
Soya Peptone	0.3 g
Proteose Peptone	0.5 g
Digested Serum	1.0 g
Yeast Extract	0.25 g
Meat Extract	0.22 g
Liver Extract	0.12 g
Dextrose	0.05 g
Soluble Starch	0.5 g
L-Tryptophan	0.02 g
L-Cysteine Hydrochloride	0.03 g
Sodium Thioglycollate	0.03 g
L-Arginine	0.1 g
Vitamin K1	0.5 mg
Hemin	1 mg
Potassium Dihydrogen Phosphate	0.25 g
Sodium Chloride	0.3 g
Agar/Gellan	15g

Wolfe's Mineral Solution (ATCC)
Add nitrilotriacetic acid to approximately 500 ml of water and adjust to pH 6.5 with KOH to dissolve the compound. Bring volume to 1.0 L with remaining water and add remaining compounds one at a time.

Nitrilotriacetic acid	1.5 g
MgSO4 . 7H2O	3.0 g
MnSO4 . H2O	0.5 g
NaCl	1.0 g
FeSO4 . 7H2O	0.1 g
CoCl2 . 6H2O	0.1 g
CaCl2	0.1 g
ZnSO4 . 7H2O	0.1 g
CuSO4 . 5H2O	0.01 g
AlK(SO4)2 . 12H2O	0.01 g
H3BO3	0.01 g
Na2MoO4 . 2H2O	.0.01 g
Distilled water	1.0 L

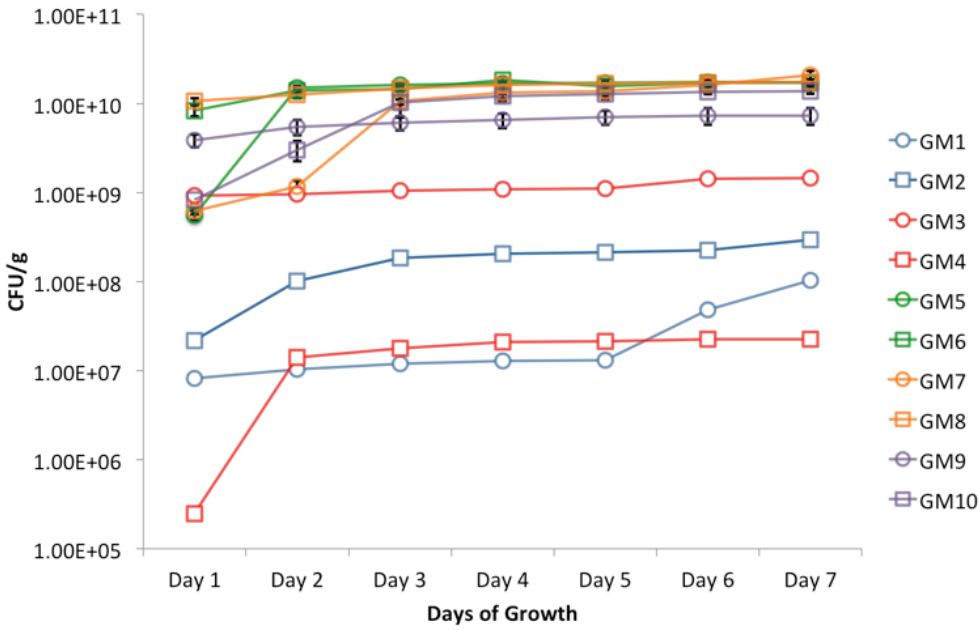
6.11 Supplementary Information

Supplementary Table 6.1 | A list of the antibiotics used in this study and their drug classes

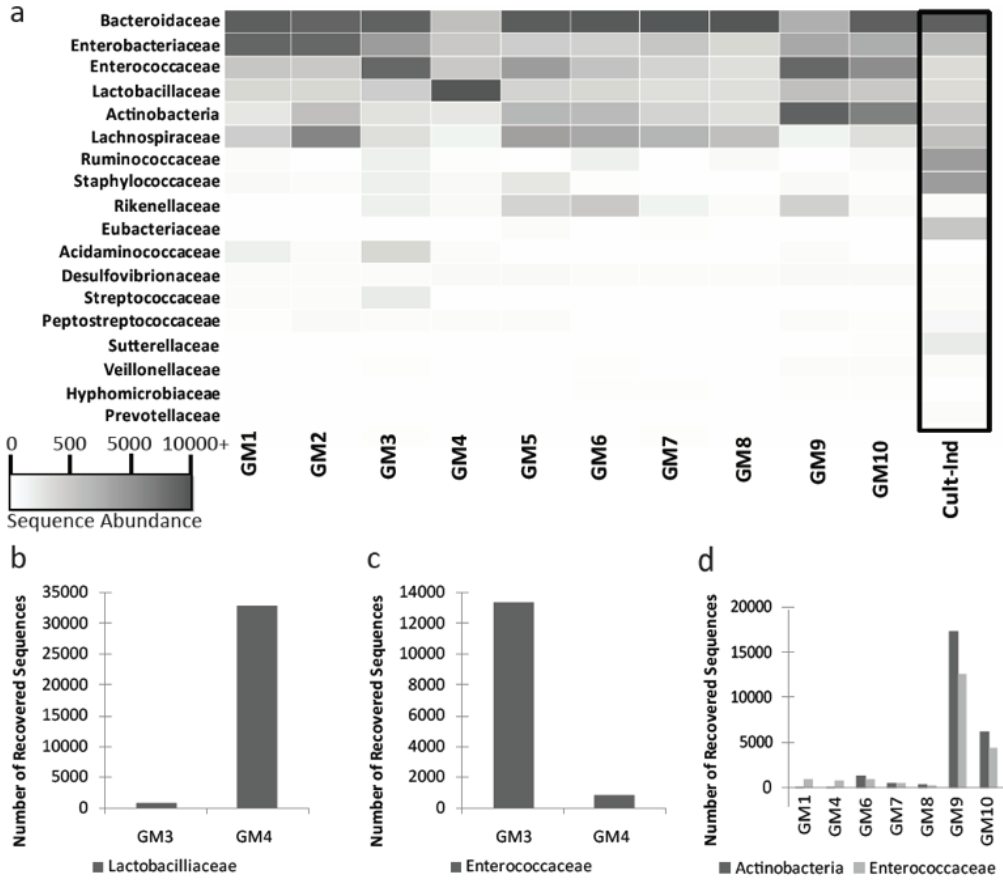
Antibiotic	Drug Class
Ampicillin (Amp)	Penicillin (Broad-spectrum)(Beta-lactamase)
Cefepime (Cefe)	Cephalosporin (4th Gen.) (Beta-lactamase)
Cefotaxime (Cefo)	Cephalosporin (3rd Gen.) (Beta-lactamase)
Cephalexin (Ceph)	Cephalosporin (1st Gen.) (Beta-lactamase)
Cefuroxime (Cefu)	Cephalosporin (2nd Gen.) (Beta-lactamase)
Chloramphenicol (Clm)	None (Bacteriostatic, Broad-spectrum)
Ciprofloxacin (Cip)	Fluoroquinolone
Clindamycin (Clinda)	Lincosamide
Dicloxacillin (Diclox)	Penicillin (Narrow-spectrum) (Beta-lactamase)
Erythromycin (Ery)	Macrolide
Gentamycin (Genta)	Aminoglycoside
Metronidazole (Metro)	Nitromidazole
Piperacillin (Pip)	Penicillin (Extended-spectrum)(Beta-lactamase)
Sulfamethoxazole (Sulfa)	Sulfonamide
Tetracycline (Tet)	Tetracycline
Vancomycin (Vanc)	Glycopeptide

Supplementary Table 6.2 | The percentage of the total library that each of the 4 sequenced genomes represented in the uncultivated samples that they were isolated from

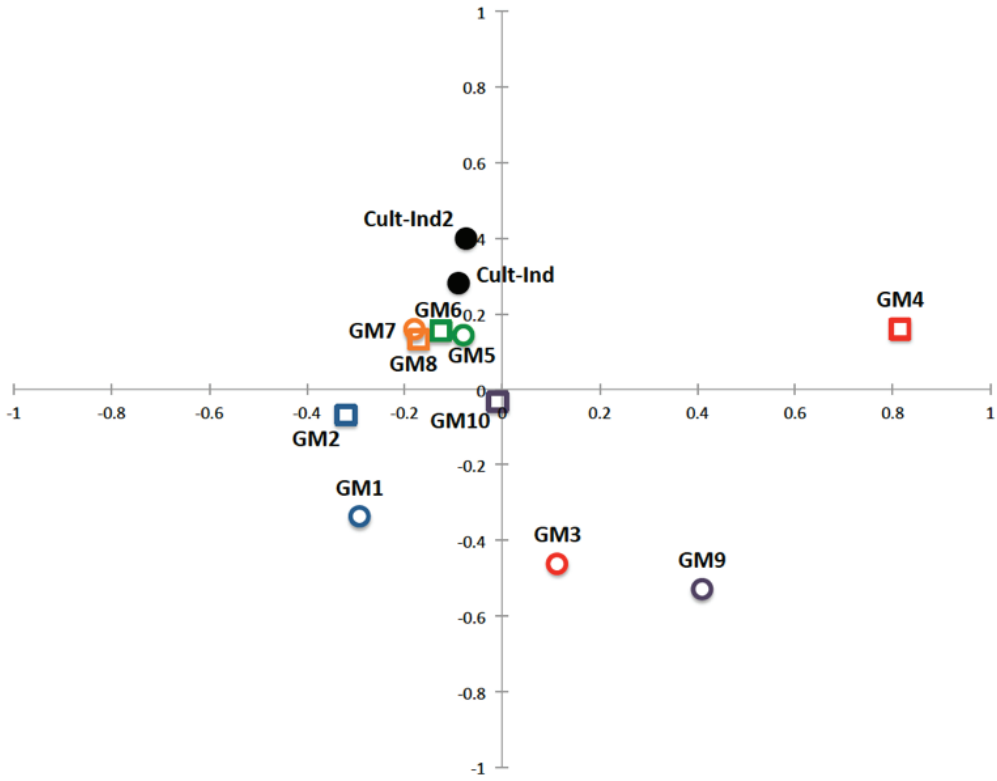
Sequenced Genome Isolate	Represented % in the Uncultivated Libraries
P1C2	0.01%
P1C11	0.08%
P1G4	Not Detected
P2C1	0.17%



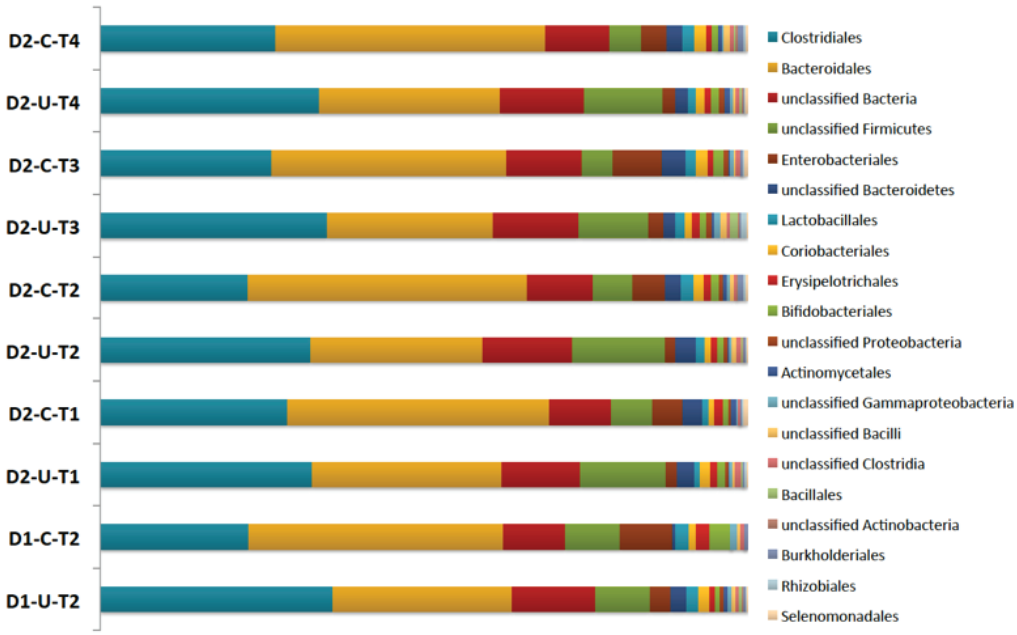
Supplementary Figure 6.1 | Plate counts of bacteria on various media. Plate counts of 10 tested media collected daily over 7 days of anaerobic growth at 37°C (These counts represent a single subject and sample (D1-T1)). The lines are the average of 5 replicates with standard error of mean (s.e.m.). Some points have no discernable error bars due to low replicate variability. Circles represent agar as a solidifying agent while squares indicate gellan as a solidifying agent.



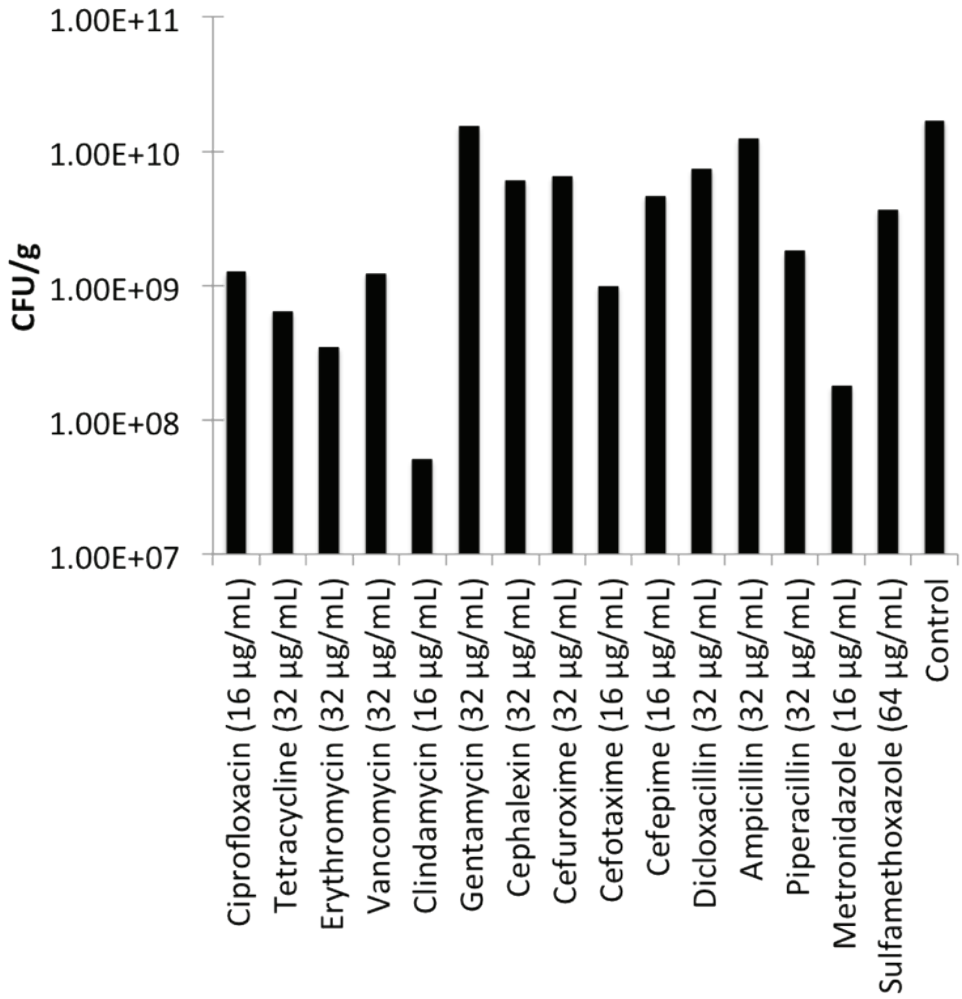
Supplementary Figure 6.2 | Taxonomic identity and relative abundance of recovered bacteria. (a) The heat map shows a family-level taxonomic comparison of the media based on relative sequence abundance and the culture-independent sample (Cult-Ind) (from Sample D1-T1). Bacteria unclassified at a family level or represented by less than 10 sequences are not shown. The order of the taxa is by total abundance of the taxonomic group over all samples. The abundance is shown on a square root scale. (b-d) Examples of taxonomic groups that were enriched for by a solidifying agent (b) Lactobacilliaceae (GM4 (gellan)), (c) Enterococcaceae (GM3 (agar)) or by media nutritional composition (d) Actinobacteria and Enterococcaceae (GM9 and GM10).



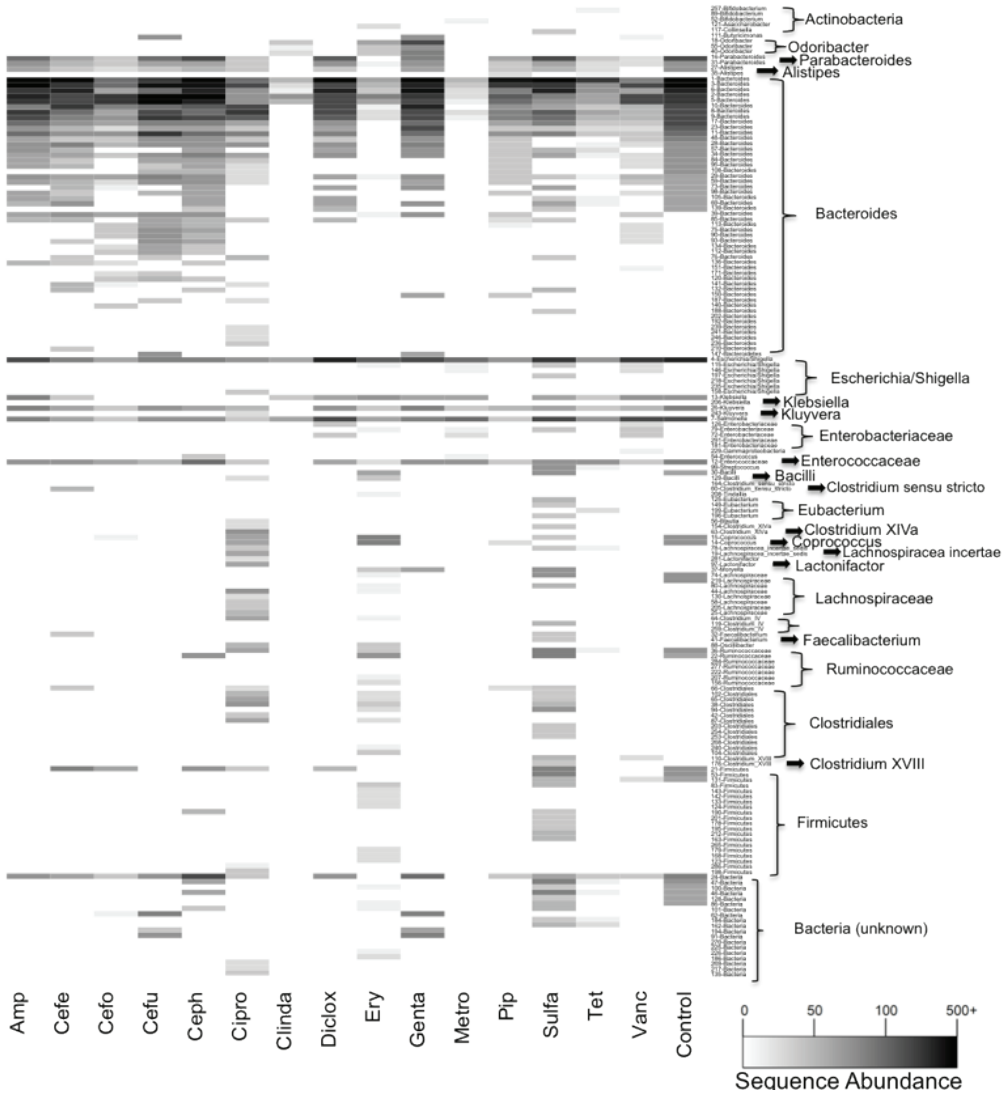
Supplementary Figure 6.3 | Nonmetric Multidimensional Scaling of Morita-Horn Indices. A Nonmetric Multidimensional Scaling (NMDS) graph calculated in mothur comparing the relatedness of all the tested culture media (Sample D1-T1 (GM1-10)) and the culture-independent sample (Cult-Ind). A replicate of the culture-independent sample (Cult-Ind2) reflects the variability in independent extractions (biological replicate) from the same sample. The R-squared for the configuration was 0.96 and the stress was 0.10 for 2 dimensions. Unfilled circles represent media containing agar while squares represent gellan-containing media.



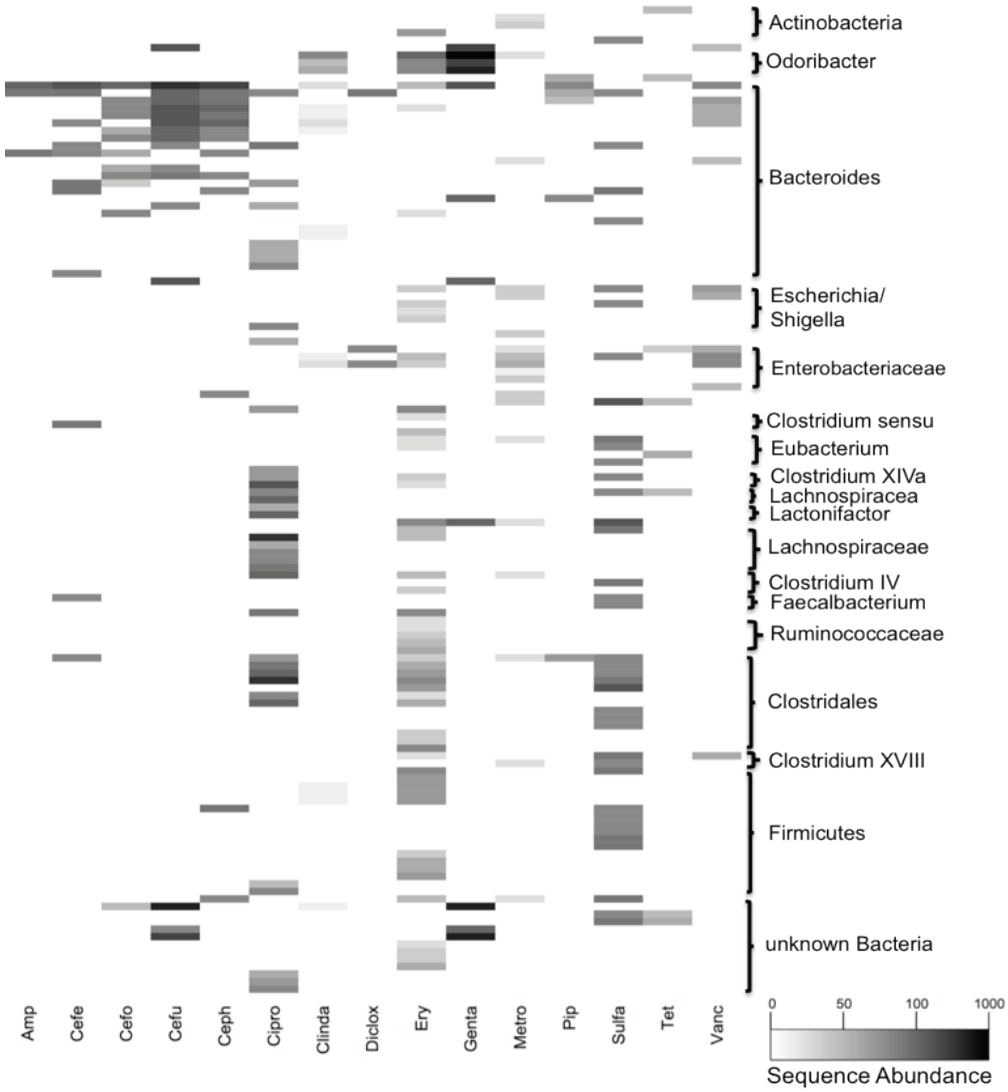
Supplementary Figure 6.4 | Repeatability of Culture Media Recovery. An order-level taxonomic comparison showing the relatedness of several samples cultured on GM7 media and their corresponding culture-independent samples (D indicates the donor, C or U indicates cultured or uncultivated, and T indicates sample number) based on relative tag abundance. Bacterial taxa represented by less than 5 sequences for all samples are not represented on the graph.



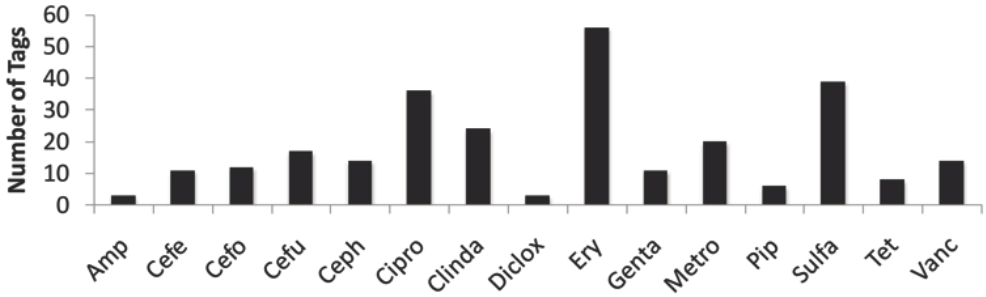
Supplementary Figure 6.5 | Antibiotic tolerance phenotyping plate counts. The total number of bacteria recovered on each of the antibiotic plates and the control plates. These values are the average of 2 replicate plates. The concentration of most antibiotics in the media was determined by using Eucast to determine MIC concentrations for bacteria recovered from the human gut. A concentration was chosen which had a 2x minimum inhibitory concentration (MIC) for the majority of commensal organisms (excluding high intrinsically resistant organisms).



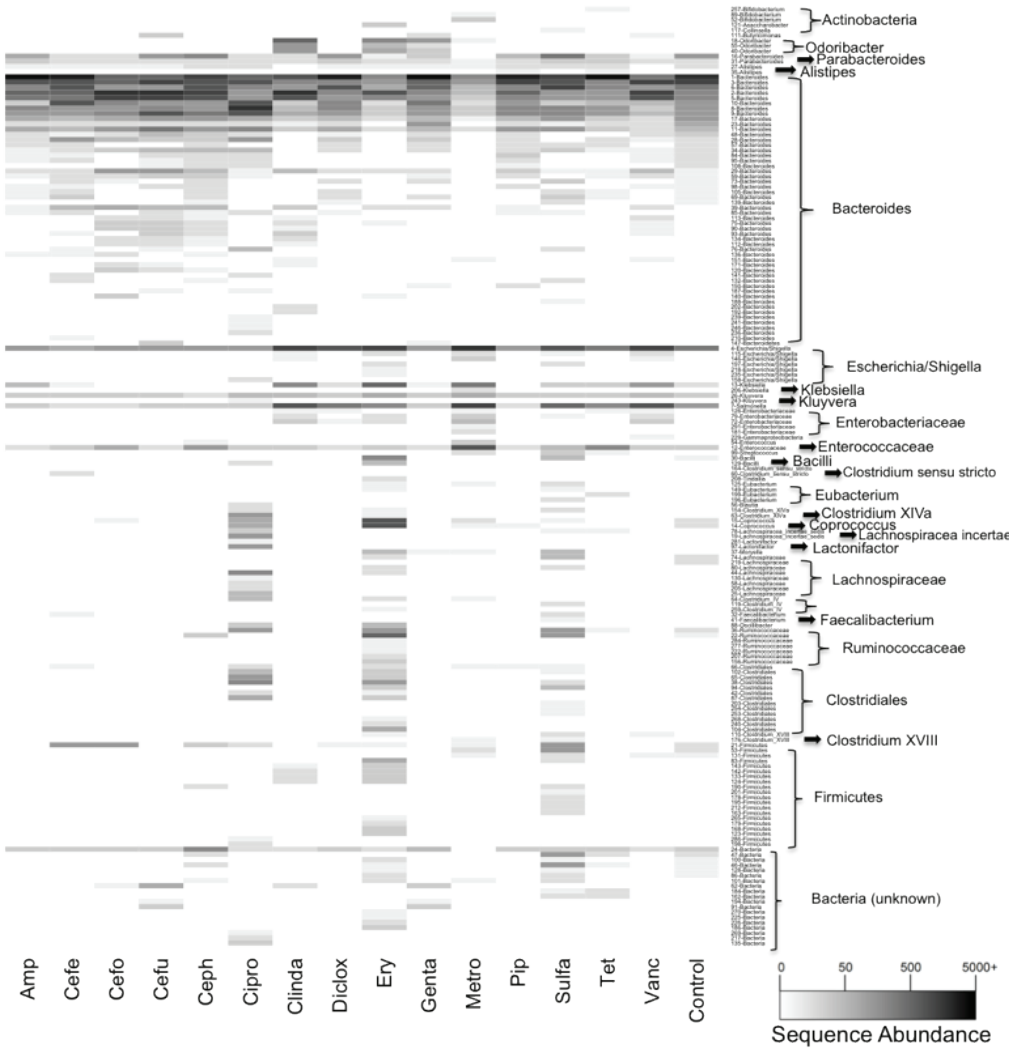
Supplementary Figure 6.6 | Total antibiotic tolerance phenotyping heat map as adjusted by plate count recovery. Heat map of the tags recovered from GM7 containing antibiotics and plates with no antibiotics (control). The heat map is sorted by taxonomic classifications and increasing number of tags is indicated by darker shades of gray. The highest level of classification was reported. Tag abundances were adjusted based on the recovery of bacteria from plate counts to more accurately reflect the comparable diversity from each antibiotic. Antibiotic abbreviations are available in Supplementary Table 6.1.



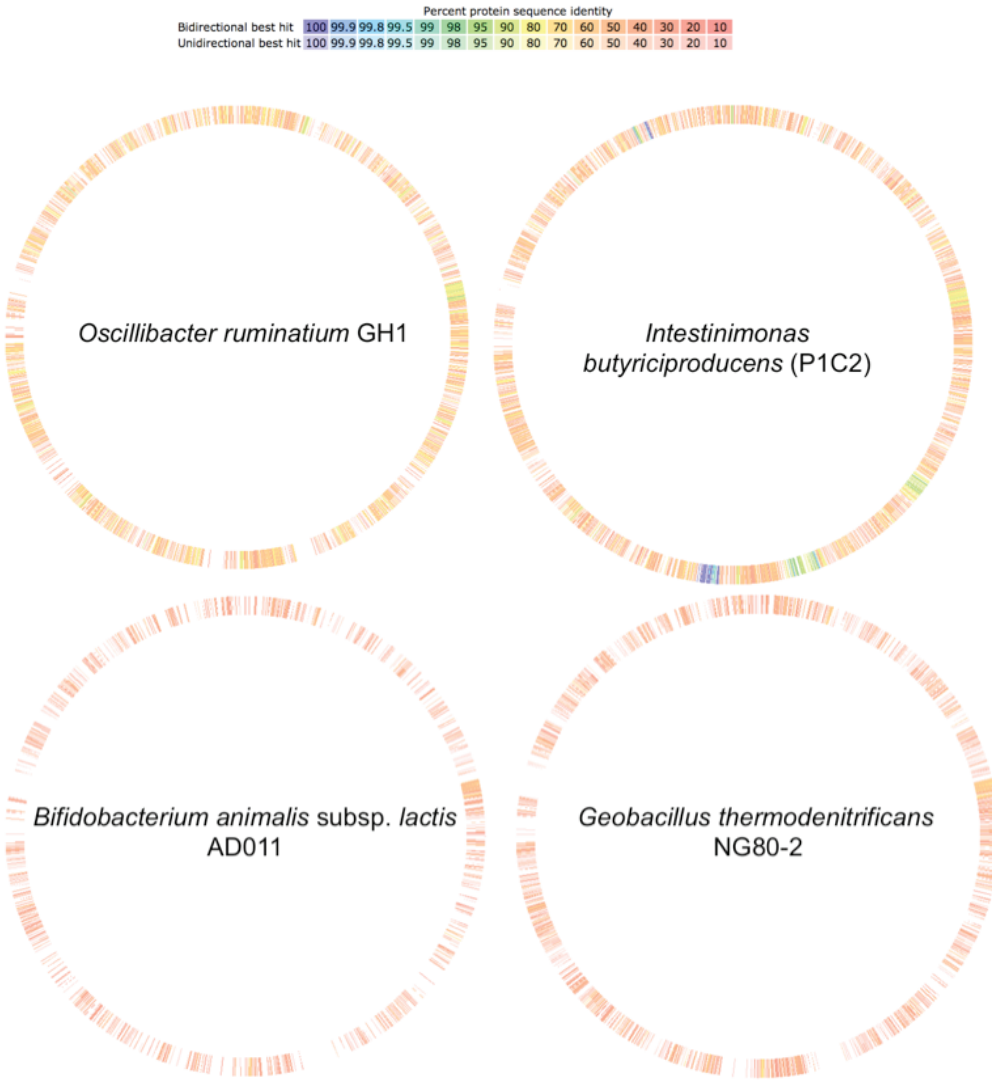
Supplementary Figure 6.7 | Antibiotic tolerance distributions in minority gut bacterial populations. Heat map of the tags making up the minority of cells within the gut community (tags not found in control). The heat map is sorted by taxonomic classifications and increasing number of tags is indicated by darker shades of gray. The highest level of classification was reported. Tag abundances were calculated by proportion of sequences recovered by each antibiotic or control sample. Antibiotic abbreviations are available in Supplementary Table 6.1.



Supplementary Figure 6.8 | Recovered tags in the minority gut bacterial community. The numbers of tags recovered for each antibiotic in the minority gut bacterial communities (those not recovered on control plates). Antibiotic abbreviations are available in Supplementary Table 6.1.



Supplementary Figure 6.9 | Total antibiotic tolerance phenotyping heat map. Heat map of the tags recovered from GM7 containing antibiotics and plates with no antibiotics (control). The heat map is sorted by taxonomic classifications and increasing number of tags is indicated by darker shades of gray. The highest level of classification was reported. Tag abundances were calculated by proportion of sequences recovered by each antibiotic or control sample. Antibiotic abbreviations are available in Supplementary Table 6.1.



Supplementary Figure 6.10 | Genome Similarity Comparisons. A comparison of the similarity of 4 publically available genomes to the sequenced genome of our P2C1 (Oscillibacter-like) isolate. The similarity calculations and mappings were generated in RAST.

CHAPTER 7

Conclusion and Perspectives

In this thesis, the genome dynamics of co-existing *Eschericia coli* lineages *in situ* of the infant gut were investigated. In the first set of isolates, where the infant was not exposed to antibiotics, we discovered that a conjugative plasmid harbouring antibiotic resistance genes was transferred from one *E. coli* lineage to another *E. coli* lineage. The transconjugant remained in the gut for months. This is, to the best of our knowledge, the first documented case of antibiotic resistance genes transferred *in situ* of the gut in a non-artifical experimental set-up and without antibiotic pressure. While the gut has been considered a hub for horizontal gene exchange, this is one of the first pieces of evidence that demonstrates that antibiotic resistance genes are transferred in the gut naturally without antibiotic pressure. In the second set of isolates, where the infant was undergoing antibiotic treatment, we genomically confirmed that transfer of an antibiotic resistance plasmid occurred between two co-existing *E. coli* lineages and we went further on to characterize the plasmid. Both of these reports highlight the role of the gut as a hub for horizontal gene exchange.

Via full genome sequencing, we identified that the lineages in the gut are highly dynamic. By comparing the *E. coli* isolates collected over the first year of the two infants' lives, we observed multiple phage infections, phage curing, large genomic deletions and plasmid acquisition and loss. Future *in vivo* studies will relate these identified genomic events to gut colonization fitness. For example, in the lineage that received the antibiotic resistance plasmid in absence of antibiotic treatment, are the isolates with the pHUSEC41-1-like plasmid better at colonizing the gut than without the plasmid? Additionally, are the isolates with the plasmid better or poorer colonizers of the gut after the large genomic deletion? By addressing these questions, we will gain a better understanding of the ability of the transconjugant to remain and to adapt in the gut.

In each of these studies, plasmids were identified with strikingly high sequence similarity to previously sequenced clinically relevant plasmids. In the sets of isolates from Chapter 4, it was the pHUSEC41-1-like plasmid carrying antibiotic resistance genes that was transferred between lineages that shared high sequence similarity to previously sequenced plasmids. The plasmid was identified in *E. coli* isolates from two geographically separated patients with hemolytic-uremic syndrome (HUS) [1, 2]. Both of these clinical *E. coli* isolates were of the serotype O104:H4, while in our study it was found in *E. coli* isolates with serotypes O73: K-:18 and O179: K12:H-. In the set of isolates from Chapter 5, where the infant was treated with antibiotics due to a urinary tract infection (UTI), we identified a pUTI89-like plasmid in the uropathogenic *E. coli* (UPEC) strain. This plasmid has been shown to play a role in the early stages of UTI [3]. This plasmid has been identified in a variety of different hosts, including those of the serotypes O18:H31 and O6:H31 [4, 5, 6]. Furthermore, this plasmid has not only been found in UPEC isolates, but also in an adherent-invasive *E. coli* strain isolated from the ileum of a Crohn's disease patient [7], and from a neonatal meningitis *E. coli* isolate [5]. Both the pHUSEC41-1-like and pUTI89-like plasmids identified in these studies highlight how clinically relevant plasmids can disseminate both with respect to host and to geographical location. In future *in vivo* studies, we would like to examine the influence of the pUTI89-like plasmid on the host's ability to colonize the gut. As this plasmid has primarily been identified in UPEC strains, and as the gut microbiota is the source of most UTIs [8], we hypothesize that this plasmid must also provide a benefit for the host to colonize the gut. Linking fitness benefits of gut colonization to this virulence-implicated plasmid may give insight as to why this plasmid is globally disseminated.

Lastly, work in this thesis also highlights that a large fraction of the gut microbiota are indeed cultivable. We demonstrated this by first assessing the antibiotic tolerance profile of the gut microbiota via V6 region 16S rRNA sequencing, and then tailoring antibiotic combinations to target previously unsequenced bacteria. One of the isolates that we cultivated and genome sequenced belonged to the genus *Oscillibactor*. Uncultivated *Oscillibactor* strains have been negatively correlated with Crohn's disease [9, 10], so the ability to cultivate these strains will provide physiological insight into these strains and will contribute to our understanding of gastrointestinal disorders and devising treatment strategies. This work also opens up possibilities to cultivate novel species that could have various health benefits and probiotic applications.

7.1 References

[1] C Kunne, A Billion, S E Mshana, J Schmiedel, E Domann, H Hossain, T Hain, C Imirzalioglu, and T Chakraborty. Complete Sequences of Plasmids from the Hemolytic-Uremic Syndrome-Associated Escherichia coli Strain HUSEC41. *Journal of bacteriology*, 194(2):532–533, December 2011.

[2] Yonatan H Grad, Paul Godfrey, Gustavo C Cerquiera, Patricia Mariani-Kurkdjian, Malika Gouali, Edouard Bingen, Terrence P Shea, Brian J Haas, Allison Griggs, Sarah Young, Qiandong Zeng, Marc Lipsitch, Matthew K Waldor, François-Xavier Weill, Jennifer R Wortman, and William P Hanage. Comparative genomics of recent Shiga toxin-producing Escherichia coli O104:H4: short-term evolution of an emerging pathogen. *mBio*, 4(1):e00452–12, 2013.

[3] C K Cusumano, C S Hung, S L Chen, and S J Hultgren. Virulence Plasmid Harbored by Uropathogenic Escherichia coli Functions in Acute Stages of Pathogenesis. *Infection and immunity*, 78(4):1457–1467, March 2010.

[4] Swaine L Chen, Chia-Seui Hung, Jian Xu, Christopher S Reigstad, Vincent Magrini, Aniko Sabo, Darin Blasiar, Tamberlyn Bieri, Rekha R Meyer, and Philip Ozersky. Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a comparative genomics approach. *Proceedings of the National Academy of Sciences of the United States of America*, 103(15):5977–5982, 2006.

[5] Dona Wijetunge, Kurundu Karunathilake, Atul Chaudhari, Robab Katani, Edward G Dudley, Vivek Kapur, Chitrita Debroy, and Subhashinie Kariyawasam. Complete nucleotide sequence of pRS218, a large virulence plasmid that augments pathogenic potential of meningitis-associated. *BMC Microbiology*, 14(1):203, August 2014.

[6] Chitrita Debroy, Mandeep S Sidhu, Upal Sarker, Bhushan M Jayarao, Adam L Stell, Nathan P Bell, and Timothy J Johnson. Complete sequence of pEC14_114, a highly conserved IncFIB/FIIA plasmid associated with uropathogenic Escherichia coli cystitis strains. *Plasmid*, 63(1):53–60, January 2010.

[7] D O Krause, A C Little, S E Dowd, and C N Bernstein. Complete Genome Sequence of Adherent Invasive Escherichia coli UM146 Isolated from Ileal Crohn's Disease Biopsy Tissue. *Journal of bacteriology*, 193(2):583–583, December 2010.

[8] Evgeni V Sokurenko, Richard Gomulkiewicz, and Daniel E Dykhuizen. Source-sink dynamics of virulence evolution. *Nature Reviews Microbiology*, 4(7):548–555, July 2006.

[9] Alan W Walker, Jennifer Ince, Sylvia H Duncan, Lucy M Webster, Grietje Holtrop, Xiaolei Ze, David Brown, Mark D Stares, Paul Scott, Aurore Bergerat, Petra Louis, Freda McIntosh, Alexandra M Johnstone, Gerald E Lobley, Julian Parkhill, and Harry J Flint. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *The ISME Journal*, 5(2):220–230, August 2010.

[10] S Mondot, S Kang, J P Furet, D Aguirre de Carcer, C McSweeney, M Morrison, P Marteau, J Doré, and M Leclerc. Highlighting new phylogenetic specificities of Crohn's disease microbiota. *Inflammatory Bowel Diseases*, 17(1):185–192, January 2011.

{2.7182818284}ϑφερτυθιοπσδφγηξκλ